ABSTRACT

WANG, KYE WON. Development of Computer Simulation Models for Drug Delivery Applications. (Under the direction of Dr. Carol K. Hall).

We developed coarse grained and implicit solvent models of DNA and lipids for the purpose of studying hydrogels and lipid bilayers. Those models are designed for use in discontinuous molecular dynamics (DMD) simulations. The DNA model is used to study the spontaneous formation of DNA-mediated hydrogels. The lipid model is used to investigate the structural properties of lipid bilayers. Our research in self-assembled hydrogels and lipid bilayers is intended to help identify their potential for use as drug delivery vehicles.

First, we develop a coarse grained nucleotide model for simulating large-scale aptamer-based hydrogel network formation. In the model, each nucleotide is represented by a single interaction site containing sugar, phosphate, and base. Discontinuous molecular dynamics (DMD) simulations are performed to simulate formation and denaturation of oligonucleotide duplexes as a function of temperature. The simulated melting temperatures of oligonucleotide duplexes are calculated in simulations of systems with different sequences, lengths, and concentrations of oligonucleotides, and compared to data from the OligoAnalyzer tool. The denaturation of oligonucleotide triplexes containing a hybridized structure of three different oligonucleotides is analyzed using both simulations and experiments. The nucleotide model is found to be a good predictor of the oligonucleotide’s hybridized state for both duplexes and triplexes. This coarse grained model has wide ranging applications in the development or optimization of DNA-based technologies including DNA origami, DNA-enabled hydrogels, and DNA-based biosensors.

Second, we present the results of discontinuous molecular dynamics (DMD) simulations aimed at understanding the formation of DNA-mediated hydrogels and assessing their drug
loading ability. Poly(ethyleneglycol) (PEG) precursors of 4 and 6 arms that are covalently
functionalized on all ends with oligonucleotides are crosslinked by a single oligonucleotide
whose sequence is complementary to the oligonucleotide conjugated to the precursor. We show
that the precursors with large molecular weight and many arms are advantageous in forming a
three-dimensional percolated network. Analysis of the percolated networks shows that the pore
diameter distribution becomes narrower as the precursor concentration, the number of arms, and
the molecular weight increase. The pore throat diameter, the size of the largest molecule that can
travel through the hydrogel networks without being trapped, is determined. The percolated
network slows the movement of molecules inside the pores. Molecules larger than the pore throat
diameter have more restrictions on their movement in the percolated network than those with
smaller sizes.

Next, we suggest an improved intermediate resolution implicit solvent model for lipids,
designed for use with discontinuous molecular dynamics (DMD) simulations. The model
improves upon the original LIME (Lipid Intermediate Resolution Model) so as to more
accurately represent the formation of lipid bilayers. The model lipid is DSPE; it has 16 coarse-
grained sites that are classified as types 1 through 6. As in the original LIME, the parameters for
the connectivity and stiffness are extracted from the explicit-solvent atomistic simulations of
lipids. However unlike the original LIME, the improved model expresses intermolecular
interactions in terms of multiple square wells. In addition, a multi-state Iterative Boltzmann
Inversion scheme is used to find the interaction parameters, so that a single set of interaction
parameters between coarse-grained sites can be used to represent the lipid bilayers at different
temperatures. The CG lipid bilayer formed quantitatively reproduces the bilayer thickness, and
produces lipid structures that compare well with atomistic simulations results.
© Copyright 2018 by Kye Won Wang

All Rights Reserved

by
Kye Won Wang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2018

APPROVED BY:

_______________________________  ________________________________
Dr. Carol K. Hall                Dr. Saad Khan
Committee Chair

_______________________________  ________________________________
Dr. Jan Genzer                  Dr. Erik Santiso
BIOGRAPHY

Kye Won Wang was born in 1982 in Seoul, South Korea. He is the son of Jin Hong Wang and Young Jae Kim, and has two younger sisters, Hyerim Wang and Hyeun Wang. He married Jean Chung in 2017 and has a daughter named Jaehee Wang. He attended Yonsei University in Seoul, Republic of Korea and earned his Bachelor of Science in Chemical Engineering in August 2008. After receiving his B.S. degree, Kye Won joined Dr. Do Hyun Kim’s lab in Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea and obtained his Master of Science degree by submitting a thesis titled “Preparation of micron-sized copolymer particles encapsulating inorganic materials by suspension polymerization in a microfluidic device”. In 2012, Kye Won came to North Carolina State University to pursue his Ph.D. in Chemical and Biomolecular Engineering. His thesis work is under the direction of Professor Carol K. Hall.
ACKNOWLEDGMENTS

I would like to express my gratitude to all of my committee members for my dissertation: Professor Saad Khan, Jan Genzer, and Erik Santiso.

I would like to thank my advisor, Dr. Carol K. Hall, for her guidance and discussions about my thesis project. Her enthusiastic dedication to research has helped me to grow as a scientist and engineer. Her most memorable teaching is that researchers must always look at all things with intellectual curiosity. Dr. Hall is not only a great scientist, she is also a great mentor. When I was stressed out, I was helped to overcome these feelings thanks to her warm advice. I greatly appreciate her efforts to teach me, and look forward to many discussions that we will have in the future.

I would like to thank all of the Hall group members that I currently work with and that I have worked with in the past. Their knowledge and enthusiasm for research has made our lab environment an extremely pleasurable place to work. I acknowledge their productive discussions with me and their assistance whenever I encountered roadblocks. I would like to especially thank the following group members: Dr. Emily Curtis and Dr. Steven Benner, for helping me get started with research; Dr. David Rutkowski, for teaching me the basics about discontinuous molecular dynamics simulations and writing computer code; Yiming Wang, for our endless brainstorming discussions. I would also like thank the post docs Drs. Xingqing Xiao and Qing Shao for their expertise and guidance. A final thanks to the younger group members: Ryan Maloney and Amelia Chen.

I would like to thank our collaborator at Texas State University, Dr. Tania Bentancourt, for many insightful discussions. I also would like to thank Joonmoo Huh for helping me to understand various computation algorithms.
I would like to thank my parents and my sisters for their support, both materially and emotionally throughout my journey toward this thesis, and throughout my entire life. Their support has been the driving force for me to move forward. I also would like to thank my wife, Jean Chung, for all always being there to support me in whatever I do. It is the best luck of my life that I met her. Finally, to my daughter Jaehee, I promise that I will always do my best to be a great dad. Thank you so much for coming to Mom and Dad.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... vii
LIST OF FIGURES .......................................................................................................... viii

Chapter 1: Introduction ................................................................................................. 1

1.1 Motivation .................................................................................................................. 1

1.2 Overview ................................................................................................................... 3

1.3 Reference .................................................................................................................. 5

Chapter 2: Development of a simple coarse-grained DNA model for analysis of oligonucleotide complex formation ................................................................................. 9

2.1 Introduction ................................................................................................................. 9

2.2 Materials and Methods ............................................................................................. 14

2.3 Results and Discussion ............................................................................................. 24

2.4 Conclusion ............................................................................................................... 30

2.5 Acknowledgements .................................................................................................. 34

2.6 References ............................................................................................................... 35

Chapter 3: Computational study of DNA-crosslinked hydrogel formation for drug delivery applications ............................................................................................................ 60

3.1 Introduction ................................................................................................................. 60

3.2 Materials and Methods ............................................................................................. 64

3.3 Results and Discussions ........................................................................................... 70

3.4 Conclusion ............................................................................................................... 78

3.5 Acknowledgements .................................................................................................. 82

3.6 References ............................................................................................................... 83

Chapter 4: Development of a Coarse-Grained model of DSPE using multistate iterative Boltzmann inversion for discontinuous molecular dynamics simulations: an improved version of LIME ........................................................................................................... 104
4.1 Introduction ......................................................................................................................... 104
4.2 Materials and Methods ........................................................................................................ 108
4.3 Results and Discussion ......................................................................................................... 112
4.4 Conclusion ............................................................................................................................ 114
4.5 Acknowledgements ............................................................................................................... 115
4.6 Reference .............................................................................................................................. 116

Chapter 5: Future works ............................................................................................................. 129

5.1 Investigation of DNA-mediated hydrogel degradation when encountered with target molecules ................................................................................................................................. 129

5.2 Investigation of the polyethylene grafting on the lipid bilayer surface for the purpose of protecting liposomes ........................................................................................................... 130
LIST OF TABLES

Table 2.1  Oligonucleotide Systems Utilized in Experimental Study\textsuperscript{a} ........................................ 41
Table 2.2  Comparison of the simulation melting temperature and the simulation temperature when the actual hybridization state is 50\%. .................................................. 42
Table 2.3  Simulated and experimental melting temperatures of System 1 to 8.
  * Experimentally determined Tm is inaccurate. See discussion within text........ 43
Table S2.1  Calculated persistence length of model oligonucleotides ................................. 44
Table S2.2  Estimated Tm for A and B Strands and Perfect Complement................................. 45
Table S2.3  The sequences of the various oligonucleotides used in the simulation and their melting temperatures in simulation temperature and in unit of Kelvin............. 46
Table 3.1  Number of CG EOs in one branch for each shape and molecular weight PEG precursor.................................................................................................................. 90
Table 3.2  The saturation degree of precursors with different shapes and molecular weights. Unit : \% .................................................................................................................. 91
Table 3.3  Diffusivity of spheres with various diameters in unpercolated (0.16 mmol/L) and three dimensionally percolated (1.55 mmol/L) networks................................. 92
Table S4.1  Minimum and maximum bond lengths for all bonded CG sites ......................... 121
Table S4.2  Interaction energy for each pair of CG type......................................................... 122
LIST OF FIGURES

Figure 2.1  Schematic illustration showing (a) the coarse grained representation of four nucleotides and (b) a pair of model complementary oligonucleotides, 5’ GGACGGTGCGAGGCG 3’ labeled with a fluorescent molecule (F) on its 3’ end, and its complement chain labeled with a quencher molecule (Q) on its 5’ end. Three types of bonds represented in (b): a covalent bond (i), pseudobond for bond angle (ii), and pseudobond for torsional angle (iii). ............................................. 47

Figure 2.2  Schematic of Fluorescent Structure-Switching DNA Probe. (A) Schematic of three-stranded DNA triplex self assembly, structure, and thermal response. (B) Chemical structure and spectra of FAM6 fluorophore and BHQ-1 quencher ........ 48

Figure 2.3  (a) Simulated fluorescence intensity as a function of simulation temperature for the oligonucleotide duplexes (sequences: 5’ CACGCCAACCTGCF 3’ and 5’ QGCAGGGTTGGCGTG 3’) at 0.1 µM. (b) Snapshot of the simulation cell showing oligonucleotide configurations in the initial (left) and final (right) states .................................................................................................................. 49

Figure 2.4  Change of fluorescence intensity (black) and unhybridized fraction (red) of F-attached oligonucleotide sequence (5’ ATACGCTGCF 3’) with simulation temperature change.......................................................................................... 50

Figure 2.5  Correlation between real temperature (K) and simulation temperature (Ts) utilized in simulations. ........................................................................................................................................ 51

Figure 2.6  Melting temperature change with concentration of oligonucleotide with Sequence of 5’ TCTGACCGF 3’ ........................................................................................................ 52

Figure 2.7  Comparison of total unhybridized fraction (red) with fluorescence intensity (black) for (a) System 1 and (b) System 3. Green and blue curves represent the unhybridized fraction of Strands A and B, respectively .............................................. 53

Figure 2.8  Comparison of fluorescence intensity between simulation and experiment for oligonucleotide System 3 .................................................................................................................................. 54

Figure S2.1 Schematics depicting real bond and pseudobonds .................................................. 55

Figure S2.2 Distribution of length of real bond and pseudo bond in CG nucleotide............... 56

Figure S2.3 Change of fluorescence intensity and unhybridized fraction of F-attached oligonucleotide sequences with reduced temperature change................................. 57

Figure S2.4 Comparison of simulated fluorescence intensity and simulated unhybridized fraction of strand A, B, and strand A+ strand B for System 1 ~ 8 ................. 58
Figure S2.5 Comparison of simulation and experiment fluorescence intensity for System 1–8.

Figure 3.1 Schematic illustrations of (a) 4-armed, (b) 6-armed precursors, and (c) a crosslinker.

Figure 3.2 The (a) initial and (b) final snapshots from a simulation containing 6-armed precursors (Blue – Precursor A, green – Precursor B) and crosslinker oligonucleotides (yellow). The inset figure of (b) is an enlarged representation of the hybridization of oligonucleotides; the base-pairing between the CG nucleotide beads is indicated by the red artificial bonds so that the hybridized state can be easily recognized. (c) The hybridized fractions of all oligonucleotides in precursors versus time. Curves of different colors indicate results obtained from different precursor concentrations.

Figure 3.3 Percolation probability in three dimensions versus the precursor concentrations for (a) 4-armed and (b) 6-armed precursor systems. Black, red, and blue curves in each figure represent precursors with molecular weight of 2, 5, and 10 kDa.

Figure 3.4 Pore diameter distribution for percolated network of (a) 6-armed/10 kDa at a variety of concentrations, (b) 6-armed/3.20 mmol/L at different molecular weights, and (c) 10 kDa / 2.46 mmol/L at different shapes of precursors.

Figure 3.5 (a) The shape of hydrogel formed by 6-armed/10 kDa at 1.55 mmol/L (blue – Precursor A, green – Precursor B, yellow – crosslinker, and red – artificial bond meaning hybridization). (b) and (c) show the positions of the pores (dark blue) in which molecules of 40.0 and 60.0 Å can move within the hydrogel (pink), respectively.

Figure 3.6 A contour map of pore of hydrogel formed by 6-armed/10 kDa at 1.55 mmol/L at x = -55.0 Å. The areas of the pores that can be reached by different sizes of molecules (20, 40, and 60 Å) are shown in blue colors with different intensities.

Figure 3.7 (a) Comparison of pore distribution and connectivity for percolated network by 6-armed/10 kDa at a precursor concentration of 1.55 mmol/L and (b) the characteristic pore diameters and pore connectivity thresholds of percolated network by 6-armed/10 kDa at various precursor concentrations.

Figure 3.8 Mean square displacement of spheres in (a) unpercolated network and (b) three dimensionally percolated networks. Each curve represents MSD of molecules of different diameters.

Figure S3.1 The projection views of (a) 3-dimensionally and (b) 2-dimensionally percolated networks with periodic duplications. Top to bottom x-y, y-z, and z-x planes. The blue box represents the simulation box.

Figure S3.2 Pore diameter distributions of networks formed by (a) 4-armed/2 kDa,
Figure S3.3 Pore connectivity threshold of percolated network by different precursors ........ 103

Figure 4.1 (a) All-atom, (b) united-atom, and (c) coarse-grained representation of DSPE. The color scheme is: blue (type1 - ethanolamine); yellow (type2 - phosphate group); red and orange (type3 and 4 - ester linkage); cyan (type5 - alkyl tail group); green (type6 - terminal tail group).............................................................. 123

Figure 4.2 Plot comparing the continuous potential energy and our discontinuous potential energy between coarse-grained site 1 and 2........................................................................ 124

Figure 4.3 (a) The convergence percentage of $g_{CG}(r)$ and $g_{A}(r)$ in SS-IBI. Comparison atomistic (black) and CG (red) RDFs between CG sites of types (b) 1-2, (b) 2-4, and (c) 3-3. ........................................................................................................ 125

Figure 4.4 The convergence percentage of $g_{CG}(r)$ and $g_{A}(r)$ in SS-IBI and MS-IBI .......... 126

Figure 4.5 Figure 5. Snapshots of a DSPE lipid bilayer in CG DMD simulation at (a) $T = 298K$, (b) $T = 340K$, and (c) $T = 370 K$. RDFs from (d) atomistic and (e) CG simulations at three temperatures ............................................................................................. 127

Figure 4.6 Comparison of the bilayer thickness measured from (a) the atomistic simulations (red), CG simulations using the (b) SS-IBI (blue) and (c) MS-IBI (black) intermolecular potentials .................................................................................. 128
CHAPTER 1
Motivation and Overview

1.1 Motivation

Drug delivery systems using nanoparticles have been developed to maximize therapeutic efficacy and minimize drug side effects by increasing residence time in vivo and selectively delivering drugs to target sites\(^1-^3\). Nanoparticles (NPs) are typically 1 to 100 nm in size\(^4\), and can be made of a variety of materials, including metals\(^5-^6\), ceramics\(^7-^8\), polymers\(^9-^{10}\), and biomolecules\(^11-^{12}\). Because of their small size, nanoparticles can pass through various barriers in the body, such as the blood-brain barrier\(^13-^{14}\). Since nanoparticles can be synthesized to have a lot of space inside of them, they can deliver large amounts of drug\(^15\). Most of the nanoparticles used in drug delivery systems are water-soluble\(^16-^{17}\), so that they can be administered intravenously, and can be discharged through normal metabolic functions without accumulating in the body\(^18-^{19}\).

In the development of nano-sized vesicles for drug delivery applications, the self assembly of bio- or biocompatible molecules has attracted attention\(^20-^{21}\). The advantages of self-assembled biomaterials are that they are simple to make and can contain a variety of functionalities such as stimuli responsive drug release\(^22-^{23}\), protection from the human immune system\(^24-^{25}\), and selective targeting to diseased areas\(^26-^{27}\). Micelles, hydrogels, liposomes, and colloidosomes are some representative structures that can be formed via self assembly\(^28-^{29}\). We are interested in DNA-mediated hydrogels and lipid bilayers because they are both biocompatible, and easy to assemble and disassemble\(^11,^{30}\). In addition, hydrogels and lipid bilayers are valuable as drug carriers because they have room for storing substances inside their structures.
In order to realize the potential of biomolecular technology, a tool that can better represent biological molecules and nanoparticles on the molecular level is required. Computer simulation is an appropriate tool for visualizing and understanding the behavior of biomolecules. Molecular dynamics (MD) is representative computer simulation tool that complements and guides experimental research. It can be used to test hypotheses about mechanisms underlying a physical phenomenon, and can be easier to perform, and less expensive, than real experiments. In addition, MD simulations allow us to observe the behavior of particles on a molecular level which is rarely possible in experiments. Although simulations are based on approximations, they are a good methodology for understanding and analyzing phenomena in nature.

Molecular dynamics simulation is divided into two types depending on the resolution: high-resolution models and low-resolution models. High-resolution models, also called atomistic models, are based on a detailed and realistic representation of molecular geometry, so that the movement of an atom is determined by the relationship with every other atom and with every solvent atom. One weakness of atomistic models is that the detailed description of molecules increases computational load and prevents applying them to large systems or long time scales. On the other hand, low-resolution models, which are also called coarse-grained models, describe the geometry and energetics of molecules in a simplified method. Coarse grained models group several atoms on a molecule into interaction sites, the molecule is then represented by a collection of interaction sites. This simplification decreases the number of interaction sites whose dynamics must be computed, improving the simulation speed and allowing access to longer time scales.

Simulations-based research on molecular self-assembly can be used to investigate design parameters for drug delivery vehicles. The work in this thesis focuses on the developments of
coarse grained models for use in discontinuous molecular dynamics simulations that can be used to study the formation of hydrogels and lipid bilayers.

1.2 Overview

In this section, we summarize Chapters 2 – 5 of this thesis. Chapter 2 is a preliminary study of DNA models for use in hydrogel simulations. Chapter 3 is a study of actual hydrogels using the model introduced in Chapter 2. Chapter 4 describes an improvement to the original coarse-grained lipid model, LIME. Finally, Chapter 5 briefly outlines future work. Each chapter includes a literature review and a bibliography.

Chapter 2 describes the development of an implicit-solvent coarse grained model for DNA. The DNA model was designed for use with discontinuous molecular dynamics simulations of DNA-mediated hydrogels. The base, sugar, and phosphate are grouped together into a single coarse grained site. The model has four distinct types of nucleotides cytosine, guanine, thymine, and adenine. The model predicts the melting temperatures of oligonucleotides with different sequences, lengths and concentrations. In addition, the denaturation of oligonucleotide triplexes containing a hybridized structure of three different oligonucleotides is captured by the DNA model; this will be used as the crosslinking moiety in hydrogel formation. The melting temperatures measured in the simulation are in agreement with the experimental data.

In Chapter 3, we simulate the formation of oligonucleotide-mediated hydrogels using the DNA model developed in Chapter 2, and assess the drug carrying ability of the hydrogel. In this work, the PEG precursors are covalently functionalized on their ends with oligonucleotides. Various shaped precursors are prepared with different physical properties (molecular weight and
number of branches of PEG). Our simulation results show that precursors with large molecular weight and many arms are advantageous in forming network structures (hydrogels). The average pore size of the network is obtained from the pore diameter distribution and the pore throat diameter is measured by pore connectivity analysis. Hard spheres of various sizes are placed in the network to calculate their mean square displacements; the drug carrying ability of the hydrogel is quantified by comparing the mean square displacements of different size spheres.

Chapter 4 provides a description of initial work performed to develop an improved version of LIME, an intermediate resolution implicit-solvent model for lipid molecules developed in the Hall group. The multiscale modeling approach used to develop the original version of LIME is modified to calculate new parameters for the improved model. The parameters for this model are obtained by collecting data from an atomistic simulation of DSPE lipid bilayers with explicit solvent. To overcome the drawbacks of LIME, multiple square wells are used to describe the intermolecular interactions between CG sites, and a multistate Boltzmann inversion scheme is applied to find the interaction energy parameters. Unlike the original model, the new model parameters can be used in simulations at different temperatures.

In Chapter 5, we describe future work on hydrogels and lipid bilayers for drug delivery.
1.3 References


7. Hodes, G., Semiconductor and ceramic nanoparticle films deposited by chemical bath deposition. *Physical Chemistry Chemical Physics* 2007, 9 (18), 2181-2196.


Chapter 2

Development of a simple coarse-grained DNA model for analysis of oligonucleotide complex formation

2.1 Introduction

Nucleic acids, including deoxyribonucleic acid (DNA), have found application in a number of areas of biological and biomedical research. They have been used for the assembly of materials for over two decades due to their ability to bind to complementary strands [1], thereby driving reversible, sequence-dependent macromolecular organization [2]. Nucleic acids have been used in the preparation of micro- and nano-scaled materials including DNA origami structures [3] and DNA hydrogels [4]. In these systems, DNA chains are used as linkers to form three-dimensional macromolecules driven by hybridization between complementary strands. Furthermore, DNA-crosslinked polymeric hydrogels have been prepared by hybridization of DNA-tethered polymer chains [5-13]. Similarly, functional nucleic acids in the form of aptamers, catalytic DNA and RNA, and aptazymes have brought about the possibility of generating intelligent materials that can respond to interactions with specific molecules [14, 15]. Aptamers are nucleic acid oligonucleotides selected through systematic evolution of ligands by exponential enrichment (SELEX) that bind molecular targets—including small molecules, proteins, and even whole cells—with high affinity in a manner similar to that through which antibodies bind to target analytes [16-18].

We are engaged in a computational and experimental research project aimed at using the binding specificity of functional biopolymers to develop bio-enabled responsive materials that could be used for drug delivery. Our long-term goals are to develop DNA-enabled micro- and nano-scaled hydrogels that can deliver therapeutic agents on-demand upon interaction with
molecular targets that are secreted in high levels by diseased tissue, and to evaluate the potential of these hydrogels as molecularly-controlled drug delivery systems. Our main hypothesis is that nucleic acids, and more specifically aptamers, can be incorporated as structural and functional components of hydrogels, using their ability to identify and interact with molecular targets as a mechanism to trigger hydrogel degradation and consequently release drugs. Interaction of the hydrogels with the target molecules is expected to result in rapid disintegration of crosslinks due to preferential binding of the aptamer to its target. When the hydrogels are formed in the presence of therapeutic agents, these agents are physically entrapped and molecularly-triggered hydrogel degradation will lead to drug release. This paper is the first in a three-part series describing development of a coarse-grained model for simulation of hydrogel formation, assessment of the hydrogel’s drug loading capacity, and hydrogel degradation when encountering target molecules.

In this paper, we introduce a coarse grained DNA model that can be used with discontinuous molecular dynamics to efficiently simulate hydrogels crosslinked with a large number of oligonucleotides. Coarse grained (CG) molecular models treat groups of several atoms as a single interaction site. Although all-atom models have been widely used to analyze the structure and thermal properties of short oligonucleotide sequences [19-21], coarse grained models have the advantage that they reduce the computational cost. Various CG models have been used to represent the structure of DNA and analyze its thermodynamic and mechanical properties [22-29]. To enable long-time observation of the hydrogel formation and target response, we use discontinuous molecular dynamics (DMD), an alternative to traditional MD in which the forces on the particles are calculated only when discontinuities in the potential are encountered [1]. DMD simulations are orders of magnitude faster than traditional MD, allowing
sampling of much wider regions of conformational space, longer time scales, and larger systems [30]. DMD has been widely used in a variety of fields requiring long-time scale simulations such as protein folding [31, 32], self assembly of biomolecules [33, 34], and structural analysis of colloidal self assembly [35]. The combination of a CG DNA model and DMD simulation allows us to analyze hybridization and denaturation of oligonucleotides. Moreover, it provides a framework for the future study of hydrogel formation, which requires simulation of large systems for long times.

The main consideration in developing a coarse grained model for DNA is that it should be capable of representing oligonucleotide hybridization and dehybridization. DNA unhybridization, also referred to as DNA denaturation, is the process by which double-stranded deoxyribonucleic acid (dsDNA) is unwound and separated into two single strands (ssDNA). The temperature at which this phenomenon is observed is called the melting temperature ($T_m$), a property that is intrinsic to each oligonucleotide complex. The percentage of GC content of DNA has a significant effect on $T_m$. The length and concentration of oligonucleotides also affect the hybridization ability. Several studies have confirmed that $T_m$ can be estimated with formulas based on the numbers of GC and AT bases [36]. In this paper, we describe the development of a DNA model that can be used for simulating hybridization of complimentary DNA strands and, ultimately, the formation of DNA-based hydrogels. Our model was designed to give a faithful estimate of the melting temperatures of oligonucleotides with different sequences at different concentrations, with predicted melting temperatures that are very close to the real melting temperature.

In addition to enabling quantitation of paired and unpaired nucleotides and visualization of the hybridization/denaturation state of oligonucleotides, the model was modified to also
enable correlation to experimentally measurable parameters. Förster resonance energy transfer (FRET) is widely used to experimentally analyze the denaturation of oligonucleotides and calculate melting temperatures. This method measures the degree of energy transfer that occurs between donor and acceptor fluorophores or quenching between a donor fluorophore and a non-fluorescent broad-band absorber molecule (quencher) that are attached to the oligonucleotides. If the donor and acceptor are in close enough proximity, energy is transferred from the donor to the acceptor, so that emission of donor fluorescence is reduced. In these mechanisms, energy transfer requires that the wavelength of the donor’s fluorescence overlaps with the wavelength range of the acceptor’s absorption, and that both donor and acceptor molecules are within close proximity (less than ~ 10 nm). In fact, the efficiency of fluorophore (F)/quencher (Q) energy transfer processes is a function of the donor/acceptor distance, $r$, as detailed by Equation 1:

$$E_T = \frac{R_0^6}{R_0^6 + r^6} = \frac{1}{1+(r/R_0)^6}$$

(1)

where $E_T$ represents the efficiency of energy transfer and $R_0$ represents the Förster distance, the distance at which the efficiency of energy transfer for a specific donor/acceptor pair is 50%. In our model, an F is tethered to one oligonucleotide and a Q is tethered to its complement oligonucleotide, so that the F and Q face each other when their respective strands form a duplex. Thus, the extent of oligonucleotide hybridization is reflected in the fluorescence of the donor fluorophore. This method allows us to derive information about the self-assembly and stability of our model oligonucleotides.

In this work we perform simulations to verify how well our DNA model calculates an oligonucleotide’s melting temperature. Each CG nucleotide is designed to represent a sugar, a phosphate, and a base as a single interaction site. Water is treated implicitly. Although the model sacrifices some of the details of the molecular level structure, it well represents the Watson–
Crick base pairing rule (A-T and C-G). The F and Q molecules involved in FRET are also coarse-grained to measure fluorescence intensity that would be obtained from experimental FRET analysis (F-Q method). These molecules are tethered to the end of a selected oligonucleotide and of its complementary oligonucleotide. Simulations start from an initial configuration in which oligonucleotides are hybridized to their complements (the oligonucleotide duplex) and randomly located in a simulation box. The average distances between F and Q are measured throughout the simulation to obtain the fluorescence intensity. The simulations are performed on oligonucleotides with different sequences and lengths. The fraction of oligonucleotides that is unhybridized (unhybridized fraction) is calculated and compared with the simulated fluorescence intensity to see how well the latter reflects the extent of oligonucleotide hybridization. Each oligonucleotide’s simulated melting temperature is calculated at different oligonucleotide concentrations, from 0.1 to 20.0 µM. At each concentration, the melting temperature is compared with the estimated melting temperature obtained from Integrated DNA Technologies OligoAnalyzer Tool, an internet-based software that predicts melting temperatures of oligonucleotides (www.idtdna.com). By so doing, we demonstrate that our DNA model is capable of representing the shift in the melting temperature as the oligonucleotide concentration changes. We also perform experiments to measure the fluorescence intensity of oligonucleotide triplexes formed when three different oligonucleotides hybridize. In total, eight different oligonucleotide triplex systems are prepared. The fluorescence intensities of each system are measured by using a spectrophotometer. Finally, we compare the simulated fluorescence intensity with the experimentally-obtained values for the eight oligonucleotide triplexes.

Highlights of the results are as follows. The fluorescence intensity curve as a function of temperature is constructed for the model oligonucleotides and their complements based on the
degree of quenching between F and Q in simulation. The shape of the curve is sigmoidal, and the simulation melting temperature is determined from the temperature at which the intensity is 50%. The unhybridized fraction and fluorescence intensity show very similar values at most temperatures, which means that the fluorescence intensity is a good indicator of the hybridization state of oligonucleotides. We find a linear relationship between the simulation temperature and the real temperature, allowing us to convert the simulation melting temperatures to real melting temperature. The resulting melting temperatures are in close proximity to the melting temperatures predicted using the OligoAnalyzer for oligonucleotides of length 8-16 base pairs (bp). The simulated melting temperatures increase as the oligonucleotide concentration increases, consistent with the characteristics of real oligonucleotides. Finally, the values of simulated melting temperatures for the oligonucleotide triplexes agree with the experimentally obtained melting temperatures. Thus, our DNA model shows the capability to predict fluorescence intensities and melting temperatures not only for a duplex but also for triplex structures of oligonucleotides.

2.2 Materials and Methods

2.2.1 Computational Model

2.2.1.1 Model Development

Figure 2.1 shows the coarse grained representation of oligonucleotides in our DNA model. Four different kinds of nucleotides are modeled to represent DNA: adenine (green), thymine (purple), cytosine (orange), and guanine (blue). Each nucleotide (including the phosphate, sugar, and base) is represented as a single coarse grained bead. The yellow bead represents F, which is 6-carboxyfluorescein (FAM6), and the black bead represents Q, which is Black Hole Quencher-1 (BHQ-1). The bead diameter, σ, for all nucleotides and for F and Q in
the system is 10.0 Å. Oligonucleotides are modeled as linear chains of 6 – 16 connected nucleotide beads with the given nucleotide sequence. The F bead is attached to the end of the oligonucleotide and the Q bead is attached to the end of that oligonucleotide’s complement. The nearest distance between F and Q is substituted into Equation 1 to determine the fluorescence intensity. The fluorescence intensity of the system is the average value of the intensity of all of the molecules at any given time point and temperature.

The oligonucleotide model has three kinds of bonds to maintain oligonucleotide structure: covalent bonds, pseudobonds to maintain bond angles, and pseudobonds to maintain torsional angles. In an oligonucleotide, the bonds between adjacent CG nucleotide beads are covalent bonds (Figure 2.1b-i). Pseudobonds for bond angles are invisible bonds between one CG nucleotide bead and its non-adjacent nearest neighbor (Figure 2.1b-ii). Pseudobonds for torsional angles are also invisible bonds but they are imposed between the reference CG nucleotide bead and the second-nearest non-adjacent neighbor CG nucleotide bead, i.e. the CG nucleotide bead three positions away (Figure 2.1b-iii). Bond stretching of all types between two CG nucleotide beads on the same chain is limited to minimum and maximum bond length values. Supplementary Figure 2.1 explains all types of minimum and maximum bond lengths graphically. When bonded CG nucleotide beads reach the minimum or maximum bond length, an infinite repulsion is generated so that they return to the prescribed bond length range [37, 38].

The minimum and maximum bond lengths between CG nucleotides in DMD simulation is obtained from atomistic simulation results. The bond length distributions for all types were extracted from GROningen MAchine for Chemical Simulations (GROMACS) simulations of united atom and explicit solvent models of oligonucleotide chains at 310 K (body temperature) with Assisted Model Building with Energy Refinement (AMBER). GROMACS is a molecular
dynamics package developed for simulations of biomolecules [39] and AMBER is a set of force fields for biomolecules [40]. Two GROMACS atomistic simulations were conducted. The first simulation contained one oligonucleotide with sequence 5’ GGACGGTGCGAGGCG 3’ in a 60 Å × 60 Å × 60 Å simulation box with 6,980 water molecules added to fill the box, and the other system contained the complementary sequence with the same number of water molecules. The number of sodium cations and chloride anions in the simulation box was 21 each in order to mimic 160 mM salt conditions. The temperature was maintained constant using the Berendsen thermostat with a time constant of 0.1 ps [41]. Pressure was also kept constant (1.0 bar) throughout the simulation. The trajectory of the oligonucleotide was collected for 30 ns with a time step of 0.001 ps. The centers of mass for each of the coarse-grained nucleotides on the oligonucleotide were calculated during the two simulations. The bond length distributions were obtained from these data by averaging the distances between the centers of mass of all bonded CG nucleotides, and the minimum and maximum bond lengths were extracted. The bond distribution plots are displayed in Supplementary Figure 2.2.

The extracted minimum and maximum bond lengths for the three types of bonds used to maintain the stiffness of CG oligonucleotides are as follows. CG distributions for bonds, angles, and torsional angles similar to those of atomistic simulation results can be realized by limiting the minimum and maximum distance obtained from the atomistic simulation to an appropriate range. The minimum bond and pseudobond lengths were refined by finding the smallest distance at which the bond distribution function reaches 30% of its maximum peak value, and the maximum pseudobond length was determined by finding the smallest distance (larger than the maximum peak distance) where the bond distribution function reaches 30% of its maximum peak value. Covalently-bonded CG nucleotides on the same chain were allowed to oscillate between a
minimum of 4.15 Å and a maximum of 5.22 Å. The pseudobond length for the bond angle has a minimum of 8.18 Å and a maximum of 9.82 Å, and the pseudobond length for the torsional angle has a minimum of 11.57 Å and a maximum of 14.00 Å. The bond distribution plots of CG oligonucleotide are compared with those from atomistic simulations of the oligonucleotide in Supplementary Figure 2.2. The bond length parameters in our model are the mean values for all CG nucleotides, indicating that they do not depend on the specific species. We measured the persistence length, that is the length over which the orientation of a polymer segment persists, of the CG oligonucleotides used in this study to confirm the validity of the bond parameters. The persistence length is defined as follows: 

\[ L_{ps} = \langle L \cdot l_0 \rangle / \langle l \rangle \]

where \( L \) is the end-to-end vector of a CG oligonucleotide, \( l_0 \) is the vector between the first and second CG nucleotides, and \( l \) is the average bond distance between two adjacent CG nucleotides [42].

The measured persistence lengths of the single strand and double strand CG oligonucleotides with 8, 10, 12, 14, 16 bps are shown in Supplementary Table 2.1. The persistence lengths of single strand oligonucleotide are 1.88 – 2.37 nm depending on the number of bps, which is close to the value for unstacked ssDNA calculated in the experimental work of Mills and co-workers (2.0 – 3.0 nm) [43]. The persistence length of double strand oligonucleotide is known to be approximately 50 nm for 150 base pairs [44]. Based on our simulation results, we found that the dsDNA persistence length for 150 base pairs is predicted to be around 26 nm by linear regression, which means our model represents dsDNA as being more flexible than it actually is.

The interaction between nucleotides in our CG model is represented using a square-well potential. The square well potential, \( u_{ij} \), between nucleotides \( i \) and \( j \) is defined to be:
\[
u_{ij}(r) = \begin{cases} 
\infty & 0 < r \leq \sigma_{ij} \\
-\epsilon_{ij} & \sigma_{ij} < r \leq \lambda \sigma_{ij} \\
0 & \lambda \sigma_{ij} < r
\end{cases}
\]

where \( r \) is the separation distance, \( \sigma_{ij} \) is the sphere diameter, \( \lambda \sigma_{ij} \) is the well width, and \( \epsilon_{ij} \) is the interaction energy between two nucleotides \( i \) and \( j \). The hard sphere diameter was chosen to be 10.0 Å and the well width was chosen to be 15.0 Å. These parameters were chosen so that the average width of paired nucleotides in a duplex structure would be 23-25 Å, which corresponds to the actual diameter of the B-DNA helix [45]. The values of the interaction energies between complementary nucleotides were assigned based on the number of hydrogen bonds between base pairs. It is well known that cytosine and guanine base pairs are the result of the formation of three hydrogen bonds, and adenine and thymine base pairs are the result of the formation of two hydrogen bonds. However, the actual intermolecular energy does not increase proportionally to the number of hydrogen bonds. In Stofer’s study, the ratio of the interaction energy of A-T and C-G pairs in water was 1 : 1.35 [46]. Thus, the energy values of \( \epsilon_{AT} \) and \( \epsilon_{CG} \) are set to 1.0 and 1.35, respectively. All other interactions, except A-T and C-G, are regarded as hard sphere interactions. The modeled F and Q molecules do not have any interactions between them or with other nucleotides as well. Note that since water is modeled implicitly in these simulations, the interaction potentials between nucleotides are essentially potentials of mean force.

### 2.2.1.2 Oligonucleotide Simulations

The details of the CG oligonucleotide simulations are as follows. For oligonucleotide duplex simulation, twenty oligonucleotide duplexes were randomly placed in a cubic box with the aim of determining the extent of hybridization and obtaining fluorescence intensity curves as
a function of temperature. The lengths of each side of the box were chosen to achieve the molar oligonucleotide concentrations of 0.1, 0.25, 0.5, 1.0, 5.0, 10.0, 15.0, and 20.0 µM. The calculated box lengths were 6926.0, 5103.0, 4050.0, 3214.5, 1879.5, 1492.0, 1303.5, and 1182.5 Å for the respective concentrations. In the simulation of oligonucleotide triplex, the number of triplexes was 50 and the length of the simulation box was 9397.1 Å for a concentration of 0.1 µM. The Anderson thermostat was used to keep the temperature constant; In this technique, the velocity of a bead is tuned to maintain the system’s Maxwell–Boltzmann velocity distribution for the set temperature [47]. The simulation temperature (T_s) is unitless and defined to be the trend line that matches the simulation melting temperature with the actual melting temperature of oligonucleotides.

The fluorescence intensity in simulations is related to the efficiency of energy transfer which is given in terms of the donor/acceptor distance, \( r \), as described by Equation 1. The separation distance between the centers of the CG donor and acceptor is used for the donor/acceptor distance. The Forster distance in the equation, \( R_o \), depends on the nature of the fluorophore and quencher molecules. In our work, \( R_o \) for the FAM6/BHQ-1 pair used in the experimental work (Section 2.2) was set to 55 Å [48]. Since the efficiency of energy transfer between the fluorophore and quencher means the proportion of energy that is not emitted as fluorescence, the fluorescence intensity is determined by subtracting the value of the efficiency of energy transfer from 1. The average distance between the centers of mass of the CG donor and accepter in the system is inserted into Equation 1 to gives the efficiency of energy transfer corresponding to the hybridization state of oligonucleotides, allowing the calculation of the fluorescence intensity. For example, if we assume the distance of F-Q at a certain temperature to
be 63.3 Å, the corresponding energy transfer efficiency will be 0.30 by the equation, and the fluorescence intensity will therefore be 0.70.

The fluorescence intensity values for oligonucleotide duplexes or triplexes in the DMD simulation as a function of temperature were obtained using the following protocol starting from a low temperature. (1) A simulation was conducted for up to 2 billion collisions at a fixed temperature and the average value of the fluorescence intensity was calculated. (2) The temperature was then increased by 0.03 and a new simulation was performed. The initial configuration of the new simulation was the final configuration generated from the previous simulation, and the average value of the fluorescence intensity was calculated as well. Steps (1) and (2) were repeated until the temperature at which the fluorescence intensity reached 1.0.

The estimated melting temperatures of oligonucleotide duplexes were obtained from the OligoAnalyzer Tool, and compared with the simulated melting temperatures at a given concentration. For oligonucleotide triplexes, the fluorescence intensity curves of simulation were compared with those of experiments at a fixed concentration of 0.1 µM.

2.2.2 Experimental Studies

2.2.2.1 Experimental Model – Oligonucleotide Triplexes

As a starting point for the demonstration of the validity of the model for the design of DNA complexes relevant to the preparation of molecularly responsive hydrogels, we focused on evaluation of the ability of the model to predict the behavior of an oligonucleotide triplex containing an aptamer. In this work, the aptamer-containing DNA structure-switching probes specific to adenosine that were first introduced by Nutiu and Li [49] were utilized. In their design, an aptamer is first synthesized with an oligonucleotide extension. Two additional partially complementary strands, one of which is labeled with a fluorophore on its 3’ end and the
other with a quencher on its 5’ end, are also utilized (Figure 2.2A). When complexed, the spatial proximity of the fluorophore and quencher results in an initial low level of fluorescence. An activation trigger such as interaction with a perfect complement of the aptamer strand, biorecognition of the aptamer’s target, or temperature increase above a critical threshold leads to dehybridization of the triplex, resulting in a measurable increase in the fluorescence signal of the system. In the work presented, the well-known adenosine aptamer (sequence 5’ ACCTGGGGGAGTATTGCGGAGGAAGGT 3’) was used as a model [50]. In addition, 6-carboxyfluorescein (FAM6) and Black Hole Quencher-1 (BHQ-1) were used as the fluorophore and quencher, respectively. FAM6 has absorption and fluorescence maxima at 495 nm and 520 nm, respectively, and an extinction coefficient of 20,960 L mol⁻¹ cm⁻¹. BHQ-1 has a broad absorption spectrum with a peak at 534 nm, and a molar extinction coefficient of 8,000 L mol⁻¹ cm⁻¹ [51]. The chemical structures and spectra of FAM6 and BHQ1 are provided in Figure 2.2B.

2.2.2.2 DNA Sequences

DNA oligonucleotides were obtained from Eurofins Genomics (Louisville, KY). Oligonucleotides were first dissolved in deionized water (obtained with a Millipore Direct-Q® water purification system equipped with a Biopak® polisher and autoclaved to eliminate DNAses) at a concentration of 100 μM and frozen in aliquots at -25 °C. All vials, pipet tips, and deionized water were autoclaved prior to use. Phosphate buffered saline (PBS, pH 7.4) was obtained from Kirkegaard & Perry Laboratories (KPL, Gaithersburg, MD) as a 10X concentrate. The 1X PBS solution is formulated as 10 mM sodium phosphate and 150 mM sodium chloride.

Table 1 lists the oligonucleotide sequences for the eight systems utilized in the experimental studies. Each system consists of three oligonucleotides: (1) a linker (L) strand that
contains an aptamer sequence (red) with an extension on its 5’ end, (2) a fluorescently labeled strand (green) that is complementary to the 5’ end of the L strand (blue), and (3) a quencher-modified strand that is complementary to the aptamer. FAM6 was used as the fluorophore on strand A1. FAM6 was covalently bound to the 5’-end phosphate of A1 oligonucleotide strands. BHQ1 was covalently bound to the 3’-end of oligonucleotide strands B1-B8.

Two main system types were utilized, as shown on Table 1. The triplex systems in Type 1 (systems 1 through 7) utilize Linker 1 (L1) as an aptamer-containing moiety. In this system type, the quencher strands hybridize partially to the aptamer and partially to five of the aptamer extension’s nucleotides. System 8 from Type 2 utilizes a shorter aptamer-containing strand (L2) in which all of strand B8 hybridizes directly with the aptamer’s nucleotides and not to the extension.

2.2.2.3 Estimation of Melting Temperatures

Estimated melting temperatures were obtained from Integrated DNA Technologies OligoAnalyzer utilizing a 0.1 µM oligonucleotide and 160 mM Na⁺ concentrations, matching the conditions utilized in experimental work and in simulations. Supplementary Table 2.2 lists the estimated melting temperatures (T_m) for each A and B strand with its perfect complement. All sequences were checked to ensure that there were no undesired stable self-dimers, hairpins, or hetero-dimers occurred using the OligoAnalyzer.

2.2.2.4 DNA Triplex Hybridization

Appropriate volumes of the appropriate oligonucleotide stocks (100 µM), 10X phosphate buffered saline (PBS), and deionized water were mixed under sterile conditions to make solutions in which each oligonucleotide concentration was 0.1 µM (1:1:1 molar ratio) and PBS
was at a 1X concentration. As an example, 20 μL of L1, 20 μL of A1, and 20 μL of B1, 1,000 μL of 10X PBS, and 8,940 μL of water were mixed for preparation of 10 mL of system 1. Aliquots of this mixture were then transferred into 1.5 mL microcentrifuge tubes, heated slowly and maintained at 95 °C for 10 minutes in an Eppendorf thermomixer. The mixtures were then allowed to come to room temperature in the thermomixer overnight prior to use.

2.2.2.5 Demonstration of Formation and Thermal Melting of DNA triplex

Hybridized oligonucleotide solutions were transferred into a quartz micro-cuvette. The fluorescence of the oligonucleotide complexes as a function of temperature was measured in a Varian Cary Eclipse fluorescence spectrophotometer from Agilent Technologies (Santa Clara, CA) equipped with a Cary Temperature Controller, a Temperature Probe Series II, and a Peltier 4 Position Multicell Holder accessory. Samples were heated and cooled from 5 °C to 95 °C over three cycles at a rate of 1 °C/min, pausing for 5 minutes at the highest and lowest temperature. Sample temperatures were monitored within the cuvette with a temperature probe coupled to the temperature controller. The fluorescence was recorded every 1 °C (λ<sub>Ex</sub> = 495 nm, λ<sub>Em</sub> = 520 nm).

The melting temperature of each system was determined from the fluorescence measurements. First, the raw fluorescence of each system was normalized by making the minimum fluorescence equal to zero and the maximum equal to 1. This was achieved by subtracting the minimum fluorescence from all data points and then dividing the corrected data by the maximum of the corrected values. Next, the T<sub>m</sub> was determined as the temperature at which the normalized fluorescence was equal to 0.5.
2.3 Results and Discussion

2.3.1 Melting Temperature Determination of Oligonucleotides from Simulated Fluorescence Intensity Curve

Figure 2.3a shows the simulation values of the fluorescence intensity versus the simulation temperature for the example oligonucleotide sequence: 5’CACGCCAACCCTGCF 3’ with its perfect complement at an oligonucleotide concentration of 0.1 µM. The black line represents the final fluorescence intensity curve of the oligonucleotide, which is the average of 20 runs at each temperature. The standard deviations of average fluorescence intensity at each temperature are used as error bars. Figure 2.3b shows snapshots of oligonucleotides at the beginning and end of the simulation. For illustrative purpose, a red artificial bond is placed between the two CG nucleotides that base pair so that the hybrid state of the nucleotide can be identified easily. Below the melting temperature, the oligonucleotide is likely to form a duplex structure. Since F is located close enough to the Q, the fluorescence intensity is low. As the temperature increases past the melting point, the fluorescence intensity increases because the oligonucleotides are separated from each other and therefore the distance between Q and F is larger. As expected, the general shape of the simulated fluorescence intensity curve is sigmoidal. As shown in Figure 2.3a, the simulation melting temperature of 5’ CACGCCAACCCTG 3’ and its perfect complement is 0.749, which is the temperature at which the fluorescence intensity is 50%. This analysis is applied to the simulations of various DNA hybrid sequences in the following sections.

2.3.2 Comparison of Fluorescence Intensity to Nucleotide Hybridization Extent

The simulation values of fluorescence intensity were compared to the total number of base pairs as a function of simulation temperature to see how accurately the fluorescence
intensity curve reflects the hybridized state of an oligonucleotide. Experimentally, oligonucleotide hybridization occurs on the molecular scale and is therefore difficult to monitor directly. Because of this, melting points and thermal stability of duplexed oligonucleotides can only be deduced using indirect analyses such as the fluorescence quenching (F-Q) method. Simulation enables direct monitoring of the hybridized fraction of oligonucleotide with single nucleotide resolution. The simulated hybridized fraction is the number of base pairs formed divided by the number of possible base pairings in the system at any given temperature. The simulated unhybridized fraction is simply expressed as:

\[
Unhybridized\; Fraction = 1 - \frac{\text{Number of Base Pairs Formed}}{\text{Maximum Number of Base Pairs}}
\]  \hspace{1cm} (2)

**Figure 2.4** compares the simulated fluorescence intensity curve (black) and the simulated unhybridized fraction curve (red) of the oligonucleotide (sequence : 5’ATACGTGCF 3’) and its complement at an oligonucleotide concentration of 0.1 µM. The simulation melting temperature of the oligonucleotide sequence is determined as 0.454 from the fluorescence intensity curve. The simulation temperature when the hybridized fraction reaches 50% is 0.451. As the simulation temperature increases above 0.5, the rate of change in both the fluorescence intensity and the unhybridized fraction slow down, reaching a maximum of 1.0 at approximately the simulation temperature of 0.6. Thus, we see that the simulated fluorescence intensity calculated via the F-Q simulation method is almost identical to the simulated unhybridized fraction of the oligonucleotides over a range of simulation temperature.

Four other oligonucleotides of different lengths (up to 16 bp) and random sequences were simulated to test whether the consistency between fluorescence intensity and the unhybridization fraction applies to every combination of oligonucleotide length and sequence (Supplementary Figure 2.3). **Table 2** shows a comparison of the simulation melting temperatures estimated from
the fluorescence intensity curve and the 50% unhybridized fraction temperatures. In all cases, the differences between the melting temperatures and 50% unhybridization temperatures are small; the oligonucleotide with the largest difference (5’ CCACGGTGGATCCA 3’) between the two temperatures has an absolute difference of about 0.010. Therefore, we find that the F-Q simulation method adequately reflects the unhybridized state of the oligonucleotide samples that we investigated.

### 2.3.3 Correlation of Simulation and Real Temperatures

To identify the relationship between simulation temperature and real temperature, and enable comparison of simulation results with experiment, we determined the melting temperatures of oligonucleotides with 18 different lengths and sequences via the F-Q simulation method. The oligonucleotide lengths were 8, 10, 12, 14, and 16, and the oligonucleotide sequences were randomly selected. **Supplementary Table 2.3** shows all the sequences of simulated oligonucleotides used in this study, their simulation melting temperatures obtained by the simulated F-Q method, and the predicted melting points from Integrated DNA Technologies OligoAnalyzer tool at oligonucleotide concentration of 0.1 µM and Na⁺ concentration of 160 mM. **Figure 2.5** shows a plot of the real melting temperatures versus the simulation melting temperatures of the 18 different oligonucleotides. The relationship between the real melting temperature from the OligoAnalyzer tool in Kelvin and simulation melting temperature by the F-Q method in terms of simulation temperature units (Tₛ) is found to be $T(K) = 114.95 \cdot T_s + 245.97 \, K$ (red dashed line). Thus, the calculated melting temperatures in dimensionless units are converted to real temperatures using this relationship. For example, an oligonucleotide with a sequence of 5’ CTCGCGTCTTF 3’ shows 50% fluorescence intensity at $T_s = 0.579$ in simulation, and this can be converted to the real temperature of 312.41 K. The melting
temperature predicted for this same sequence using the OligoAnalyzer tool is 313.40 K, making the gap between the two temperatures small, 0.99 K. The average difference of the real melting temperature and the converted melting temperature is of 2.44 K for all 18 oligonucleotide sequences. Thus, we conclude that the simulation temperature can be expressed by a first-order relationship to the real temperature.

### 2.3.4 Effect of Concentration on Oligonucleotide Duplex Stability

To ensure that the concentration dependence of oligonucleotides’ melting temperatures is well represented in our simulations, we measured the melting temperatures at different concentrations. One interesting property of DNA hybridization is that the stability of hybridized oligonucleotides increases with the concentration of the oligonucleotide and thereby increases their melting temperature as well. **Figure 2.6** compares simulated melting temperatures (red circles) to the predicted melting temperatures (black squares) obtained from the OligoAnalyzer for the sequence 5’ TCTGACCGF 3’ as a function of concentration. The change of simulated melting temperature with concentration agrees with the predicted melting temperature variation within the concentration range used in this study. While the melting temperature increases with the oligonucleotides’ concentration, its rate of increase gradually decreases. This tendency can be explained in terms of the frequency of intermolecular collisions between oligonucleotide strands. The higher the probability of intermolecular collision becomes, the higher the probability that unhybridized oligonucleotides will form duplexes will be. Highly concentrated oligonucleotides are more likely to join together in duplex structures than dilute systems of oligonucleotides at the same temperature, and therefore a higher concentration system needs higher temperatures than a lower concentration to reach 50% denaturation and therefore 50% increase in fluorescence.
intensity. Our DNA model is thereby able to reasonably calculate the melting temperature of oligonucleotides at different concentrations.

### 2.3.5 Comparison of Experimental and Simulated Oligonucleotide Triplex Stability and Melting Temperatures

We measured the fluorescence intensity and unhybridized fraction versus temperature of the oligonucleotide triplexes listed in Table 1 in simulations at 0.1 µM to see if the triplex fluorescence intensity reflects the hybridization state as well as it does for the duplexes. As described in Table 1, each triplex is composed of a long Linker strand (L), as well as two partially complementary strands A and B labeled with a fluorophore (F) and quencher (Q), respectively. Strand A forms a longer hybridization pair with the Linker strand, and therefore forms a more stable complex with higher melting temperature than Strand B. Between systems 1 and 7, the Linker (sequence: 5’CCTGCCACGCTCCGCTCACTGGGGGAGTATTGCGGAGGAAGGT 3’) and Strand A (sequence: 3’ GGACGGTGCGAGGCGF 5’) sequences stay constant; only strand B differs in its length: it is longest in System 1 and shortest in System 7. Thus, the melting temperature difference between Strand A and Strand B is at a minimum in System 1, and at a maximum in System 7. Figure 2.7 compares the simulated fluorescence intensity (black), and the unhybridized fractions of Strand A (blue) and Strand B (green) when hybridized to the linker for two representative cases, Systems 1 and 3. The sequence of Strand B in System 1 is 3’QGTGACTGGACCCCC 5’, and that in System 3 is 3’QGTGACTGGACCC 5’. There are two major differences between the graphs of Systems 1 and 3. First, the total unhybridized fraction curves of Strands A and B (red in Figure 2.7a) shows a normal sigmoidal profile in System 1. However, the total unhybridized fraction curve in System 3 has an extra inflection point around 335 K (red in Figure 2.7b). Another feature of System 3 that differs from
System 1 is that the simulated fluorescence intensity curve (black in Figure 2.7b) and the unhybridized fraction curve of Strand B (green in Figure 2.7b) are nearly identical. From analysis of all of the systems in Supplementary Figure 2.4 a-g, we find that Systems 3 through 7 show an inflection in the total unhybridized fraction curve and agreement between the simulated fluorescence intensity curve and Strand B’s unhybridized fraction curve. This behavior is a result of the large differences in thermal stability of Strands A and B in these systems (the melting temperature differences between Strands A and B for these systems are greater than 17.6 K). This difference causes sequential—as opposed to simultaneous—dehybridization; first Strand B dehybridizes and then Strand A dehybridizes. As Strand B leaves the triplex, FRET quenching immediately decreases, leading to the changes in fluorescence intensity associated with the hybridized state of Strand B. A similar phenomena occurs for System 8 (T_{m,A} – T_{m,B} = 17.9K), but in this case it is less pronounced than in the other systems (Supplementary Figure 2.4e). In summary, our simulations suggest that when Strands A and B composing the oligonucleotide triplex have a melting temperature difference above 17.6 K, the fluorescence intensity of the triplex is decided by the unhybridization of the oligonucleotide with the lower melting temperature. In contrast, when the melting temperature difference between the two oligonucleotides is below 17.6 K, the fluorescence intensity is the result of the simultaneous denaturation of the two oligonucleotides that join the triplex together.

We then compared the simulated fluorescence intensity values for the oligonucleotide triplexes listed in Table 1 to the fluorescence intensity values from our own experiments over a range of temperatures. Figure 8 shows a comparison between the simulation (black line) and experimental (red line) fluorescence intensities versus temperature for System 3 (Linker : 5’CCTGCCACGCTCCGCTCAGACTGACCTGGGGAGTATTGCCGGAGGAAGGT 3’, Strand A
GGACGGTGCGAGGCG 5’, and Strand B : 3’QGTGACTGGACCC 5’). The results for Systems 1–2 and 4–8 can be found in Supplementary Figure 2.5 a, b, and d - h. The simulated fluorescence intensity curve for System 3 is in reasonable agreement with the experimental fluorescence intensity curve. The simulated melting temperature of System 3 is 325.94 K, very similar to the experimental value of 326.69 K. Table 3 shows the calculated melting temperatures from simulations and the experimentally determined melting temperatures of all 8 oligonucleotide triplex systems. The average difference between the simulated and experimental melting temperatures is 2.73 K, although the largest difference is 11.8 K for System 7. It is worth mentioning that the experimental determination of the T_m for systems containing shorter oligonucleotides (Systems 5, 6, and 7) has problems due to the low temperatures required to reach complete hybridization. Below 5 °C, the required temperature range to reach hybridization stability for these systems, cooling results in significant condensation outside of the cuvettes used for fluorescence spectroscopy, and this prevents accurate data collection. This issue is especially important for System 7 which would require temperatures below 0 °C to reach 100% hybridization, something that is not experimentally possible in our setup. Except for System 7, the simulated melting temperature values determined by our model generally agree with the experimental values. We conclude that our model is good enough to allow calculation of the melting temperatures of any oligonucleotide triplex in which the experimental measurements can be made accurately.

2.4 Conclusion

We have developed a coarse-grained model of oligonucleotide strands for use with discontinuous molecular dynamics simulations. The model has 4 different nucleotide types: adenine, thymine, cytosine, and guanine. Each nucleotide is represented by a coarse grained bead
that contains a sugar, a phosphate, and a base. Each oligonucleotide strand is a chain of coarse grained nucleotides arranged in sequence order. Adenine beads are designed to interact with thymine, and cytosine beads are designed to interact with guanine to satisfy the Watson-Crick base pairing rule. The values of the interaction energies for A-T and C-G were obtained from Stofer [46]. We introduced three types of bonds to maintain oligonucleotide stiffness: covalent bonds, pseudobonds to maintain bond angles, and pseudobonds to maintain torsional angles. The distributions of the lengths of the three types of bonds used in DMD simulation were obtained from GROMACSs atomistic simulation data on oligonucleotides.

Simulations were performed to see how well our CG DNA model represents the hybridization of actual oligonucleotides (duplex form) under various conditions. First, we constructed fluorescence intensity curves of oligonucleotides with different lengths (8-16 bps) using the CG DNA model. The values of the fluorescence intensity were calculated by measuring the energy transfer efficiency between a model fluorophore donor and accepter tethered to the end of an oligonucleotide and its complement. The melting temperature was defined to be the temperature at which the simulated fluorescence intensity was 50%. The number of unhybridized pairs of oligonucleotides versus temperature was also calculated from the simulations. The temperatures at which the fluorescence intensity was 50% corresponded to that at which the unhybridization fraction was 50%. This implies that the value of the simulated fluorescence intensity represents the unhybridized fraction of oligonucleotides at a given temperature. Next, we measured the melting temperatures of oligonucleotides over a range of concentrations and compared this with the predicted melting temperatures obtained from OligoAnalyzer. The simulation results for melting temperatures at various concentrations
corresponded to their predicted values. This shows that our CG model can be used at various concentrations.

The simulated fluorescence intensity was compared with the simulated unhybridized fraction of oligonucleotide triplexes, and the simulated melting temperatures were compared with experimental melting temperatures. Eight different types of triplexes were prepared; the Strand A and Linker sequences that make up the triplexes were the same for the eight triplexes, but the last oligonucleotide (Strand B) was of a different length. A graph of the simulated fluorescence intensity of all the triplexes versus temperature shows a sigmoidal shape as in the case of the oligonucleotide duplex. Although the fluorescence intensity of the oligonucleotide duplex correctly reflected the unhybridization state between oligonucleotide and its complement, the agreement between the simulated and experimental fluorescence intensity for the oligonucleotide triplex depends on the differences between the melting temperatures of the two strands. When the melting temperature difference between Strands A and B is above 17.6 K, the fluorescence intensity of the oligonucleotide triplex reflects the unhybridization state of the oligonucleotide with the lower melting temperature due to the sequential separation of Strand A and Strand B. However, when the melting temperature difference is below 17.6 K, Strands A and B separate from the triplex simultaneously and it is not useful to interpret the fluorescence intensity in terms of the unhybridization state of a specific oligonucleotide. The experimental fluorescence intensity curve of the triplexes was sigmoidal in shape as well. From the fluorescence intensity curve of the oligonucleotide triplex, we obtained the temperatures at which the intensity was 50%, and then compared them with the values obtained experimentally. The difference between the simulated melting temperatures and the experimental values are 2.73 K on average.
This study lays the foundation for future work on analysis of hydrogels crosslinked by oligonucleotides. The CG DNA model that we have developed is capable of representing an oligonucleotide duplex’s hybridization state at different temperatures. Moreover, we have shown that our model is useful for analyzing the melting temperature of oligonucleotide triplexes which consist of strand A, Strand B, and Linker. In future research, we plan to tether Strand A to the end of one polymer chain (Precursor A), Strand B to the end of another polymer chain (Precursor B), and use a Linker oligonucleotide to crosslink those two precursors, forming a hydrogel. The fluorescence intensity analysis of the triplex will provide a basis for determining if a hydrogel can be formed at any given temperature.

Although the combination of CG DNA model with the DMD approach is able to simulate the melting temperatures of oligonucleotides, it has some limitations. First, as the coarse graining method groups the sugar, phosphate, and base into a single interaction site, it does not account for the details of the interactions at the molecular level such as hydrogen bonding and π-π stacking. Nevertheless, although our model does not give a full description of the association and dissociation behavior of DNA, it is well suited to the study of hydrogels that use DNA as a cross-linker. Since this model gives a reasonable representation of the hybridization state of DNA at specific temperatures, we believe that it is meaningful enough to be used for this purpose. Second, the DNA model was developed to model one fixed salt concentration (160 mM). To model melting temperatures at different salt concentrations, we could make the CG DNA interaction energy parameters be a function of the salt concentration, or add a correction term to the relationship between the simulation temperature and actual temperature to reflect melting temperature shift caused by cation concentration change. Third, our model does not display the helical structure that occurs when the oligonucleotide hybridizes with its complement. This
shortcoming can be remedied by making the base pair interactions anisotropic. Fourth, the DNA model is validated only for oligonucleotide systems with relatively short lengths (less than 16 bp). Finally, although model is reasonable enough to find the melting temperatures of oligonucleotide triplex, there are parts where the simulated fluorescence intensity graph does not match the experiment because the exact location of occurring the energy transfer is not reflected.

2.5 Acknowledgements

This work was supported by the National Institutes of Health (EB006006), the NSF’s Research Triangle MRSEC on Programmable Soft Matter, DMR-1121107, and the NSF Partnership for Research and Education in Materials (PREM), DMR-1205670.
2.6 References


Table 2.1. Oligonucleotide Systems Utilized in Experimental Study.a.

<table>
<thead>
<tr>
<th>System</th>
<th>No.</th>
<th>Sequencesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>L1</td>
<td>5’ CCTGCCACGCTCCGC—T—CACTGACCTGGGGGAGTATTGCGGAGGAAGGT 3’</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACTGGACCCC 5’ B7</td>
</tr>
<tr>
<td>Type 2</td>
<td>L2</td>
<td>5’ CCTGCCACGCTCCGC—T—ACCTGGGGGAGTATTGCGGAGGAAGGT 3’</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>A1 3’ GGACGGTGCGAGGGCF QGTGACCCCCCTC 5’ B8</td>
</tr>
</tbody>
</table>

a Linker strands L1 and L2 hybridize with the respective fluorophore or quencher-labeled strands A (blue) and B (green) by complementary base pairing to form systems 1 through 8.
b In each sequence, A = adenine, T = thymine, G = guanine, C = cytosine, F = FAM6, Q = BHQ1.
Table 2.2. Comparison of the simulation melting temperature and the simulation temperature when the actual hybridization state is 50%.

<table>
<thead>
<tr>
<th>Oligonucleotide Sequence</th>
<th>Melting Temperature by F-Q Method ($T_s$)</th>
<th>Temperature of 50% Hybridization State ($T_s$)</th>
<th>Relative Difference ($T_s$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCTGACCG</td>
<td>0.481 ± 0.009</td>
<td>0.478 ± 0.009</td>
<td>0.003</td>
</tr>
<tr>
<td>GTCGGCAGCA</td>
<td>0.609 ± 0.013</td>
<td>0.601 ± 0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>CACGCGCTGTGGT</td>
<td>0.664 ± 0.006</td>
<td>0.656 ± 0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>CCACGCGTGATCCA</td>
<td>0.750 ± 0.005</td>
<td>0.740 ± 0.006</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 2.3. Simulated and experimental melting temperatures of System 1 to 8.

* Experimentally determined $T_m$ is inaccurate. See discussion within text.

<table>
<thead>
<tr>
<th>System</th>
<th>Melting Temperature from Simulation (K)</th>
<th>Experimental Melting Temperature (K)</th>
<th>Difference (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>334.26 ± 1.24</td>
<td>332.24</td>
<td>2.02</td>
</tr>
<tr>
<td>2</td>
<td>330.49 ± 0.76</td>
<td>329.49</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>325.94 ± 0.79</td>
<td>326.69</td>
<td>-0.75</td>
</tr>
<tr>
<td>4</td>
<td>318.69 ± 0.51</td>
<td>320.06</td>
<td>1.36</td>
</tr>
<tr>
<td>5</td>
<td>307.67 ± 0.50</td>
<td>304.82</td>
<td>2.85</td>
</tr>
<tr>
<td>6</td>
<td>305.02 ± 3.48</td>
<td>305.51</td>
<td>-0.49</td>
</tr>
<tr>
<td>7</td>
<td>302.31 ± 0.42</td>
<td>290.51*</td>
<td>11.8</td>
</tr>
<tr>
<td>8</td>
<td>324.40 ± 1.09</td>
<td>322.87</td>
<td>1.53</td>
</tr>
</tbody>
</table>
Table S2.1. Calculated persistence length of model oligonucleotides.

<table>
<thead>
<tr>
<th>Oligonucleotide sequence</th>
<th>Persistence length (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single strand</td>
</tr>
<tr>
<td>5’TCTGACCG3’</td>
<td>1.88 ± 0.26</td>
</tr>
<tr>
<td>5’AGACACCACGC3’</td>
<td>2.07 ± 0.34</td>
</tr>
<tr>
<td>5’CCGCGGTGGATC3’</td>
<td>2.19 ± 0.45</td>
</tr>
<tr>
<td>5’CCGCGGTGGATCCA3’</td>
<td>2.28 ± 0.56</td>
</tr>
<tr>
<td>5’CCGCGGTGGATCCACC3’</td>
<td>2.37 ± 0.65</td>
</tr>
</tbody>
</table>
**Table S2.2.** Estimated Tₘ for A and B Strands and Perfect Complement.

<table>
<thead>
<tr>
<th>Strand</th>
<th>Estimated Tₘ (°C)</th>
<th>Estimated Tₘ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>65.1</td>
<td>338.25</td>
</tr>
<tr>
<td>B1</td>
<td>56.5</td>
<td>329.65</td>
</tr>
<tr>
<td>B2</td>
<td>52.3</td>
<td>325.45</td>
</tr>
<tr>
<td>B3</td>
<td>47.5</td>
<td>320.65</td>
</tr>
<tr>
<td>B4</td>
<td>42.0</td>
<td>315.15</td>
</tr>
<tr>
<td>B5</td>
<td>35.4</td>
<td>308.55</td>
</tr>
<tr>
<td>B6</td>
<td>29.0</td>
<td>302.15</td>
</tr>
<tr>
<td>B7</td>
<td>22.8</td>
<td>295.95</td>
</tr>
<tr>
<td>B8</td>
<td>47.2</td>
<td>320.35</td>
</tr>
</tbody>
</table>
**Table S2.3.** The sequences of the various oligonucleotides used in the simulation and their melting temperatures in simulation temperature and in unit of Kelvin.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Melting Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulation</td>
</tr>
<tr>
<td>1  5’TCTGACCG3’</td>
<td>0.481</td>
</tr>
<tr>
<td>2  5’ATACGTGC3’</td>
<td>0.454</td>
</tr>
<tr>
<td>3  5’TCTGCGAC3’</td>
<td>0.469</td>
</tr>
<tr>
<td>4  5’AGACCACG3’</td>
<td>0.496</td>
</tr>
<tr>
<td>5  5’AGACCACCGC3’</td>
<td>0.609</td>
</tr>
<tr>
<td>6  5’GGACTGACCC3’</td>
<td>0.581</td>
</tr>
<tr>
<td>7  5’TGCTCTAGGC3’</td>
<td>0.550</td>
</tr>
<tr>
<td>8  5’TGCCTGTCGA3’</td>
<td>0.555</td>
</tr>
<tr>
<td>9  5’CACGCCTGTTG3’</td>
<td>0.664</td>
</tr>
<tr>
<td>10 5’CCACGGAGCCTA3’</td>
<td>0.653</td>
</tr>
<tr>
<td>11 5’CGCTCGCAATTTC3’</td>
<td>0.626</td>
</tr>
<tr>
<td>12 5’CCACGCACCTA3’</td>
<td>0.663</td>
</tr>
<tr>
<td>13 5’CCGCGGTGGATCCA3’</td>
<td>0.750</td>
</tr>
<tr>
<td>14 5’ACCCGGCCCGAGGTA3’</td>
<td>0.789</td>
</tr>
<tr>
<td>15 5’CGTCCCAACCGCAC3’</td>
<td>0.754</td>
</tr>
<tr>
<td>16 5’CAGATGCCGGGACC3’</td>
<td>0.730</td>
</tr>
<tr>
<td>17 5’CCGCGGTGGATCCACC3’</td>
<td>0.841</td>
</tr>
<tr>
<td>18 5’TGCCTACCGCACCCTG3’</td>
<td>0.820</td>
</tr>
</tbody>
</table>
Figure 2.1. Schematic illustration showing (a) the coarse grained representation of four nucleotides and (b) a pair of model complementary oligonucleotides, 5’ GGACGGTGCAGGCG 3’ labeled with a fluorescent molecule (F) on its 3’ end, and its complement chain labeled with a quencher molecule (Q) on its 5’ end. Three types of bonds represented in (b): a covalent bond (i), pseudobond for bond angle (ii), and pseudobond for torsional angle (iii).
**Figure 2.2.** Schematic of Fluorescent Structure-Switching DNA Probe. (A) Schematic of three-stranded DNA triplex self assembly, structure, and thermal response. (B) Chemical structure and spectra of FAM6 fluorophore and BHQ-1 quencher.
Figure 2.3. (a) Simulated fluorescence intensity as a function of simulation temperature for the oligonucleotide duplexes (sequences: 5’ CACGCCAACCCTGCF 3’ and 5’ QGCAGGTTGGCGTG 3’) at 0.1 µM. (b) Snapshot of the simulation cell showing oligonucleotide configurations in the initial (left) and final (right) states.
Figure 2.4. Change of fluorescence intensity (black) and unhybridized fraction (red) of F-attached oligonucleotide sequence (5’ ATACGTGCF 3’) with simulation temperature change.
Figure 2.5. Correlation between real temperature (K) and simulation temperature ($T_s$) utilized in simulations.
Figure 2.6. Melting temperature change with concentration of oligonucleotide with sequence of 5’ TCTGACCGF 3’.
Figure 2.7. Comparison of total unhybridized fraction (red) with fluorescence intensity (black) for (a) System 1 and (b) System 3. Green and blue curves represent the unhybridized fraction of Strands A and B, respectively.
Figure 2.8. Comparison of fluorescence intensity between simulation and experiment for oligonucleotide System 3.
Figure S2.1. Schematics depicting real bond and pseudobonds.
Figure S2.2. Distribution of length of real bond and pseudo bond in CG nucleotide.
**Figure S2.3.** Change of fluorescence intensity and unhybridized fraction of F-attached oligonucleotide sequences with reduced temperature change.
Figure S2.4. Comparison of simulated fluorescence intensity and simulated unhybridized fraction of strand A, B, and strand A+ strand B for System 1 ~ 8.
Figure S2.5. Comparison of simulation and experiment fluorescence intensity for System 1–8.
Chapter 3

Computational study of DNA-crosslinked hydrogel formation for drug delivery applications

3.1 Introduction

Hydrogels are three-dimensional networks of polymer chains\textsuperscript{1} that are crosslinked via physical or chemical means\textsuperscript{2}. They are valued for their high water holding capacity (> 90 % water)\textsuperscript{3-4}, which gives them a flexibility that is similar to that of natural tissue and makes them biocompatible\textsuperscript{5-6}. Hydrogels are used in a variety of applications including contact lenses\textsuperscript{7}, wound healing dressings\textsuperscript{8-9}, tissue engineering\textsuperscript{10-14}, biosensors\textsuperscript{15}, and drug delivery\textsuperscript{16-18}.

The porous (or network) structure of the hydrogels makes them well suited for carrying small molecules such as therapeutics\textsuperscript{19} and biomolecules such as growth factors, hormones and protein therapeutics\textsuperscript{20-21}. The size of pores in a hydrogel controls the movement (diffusion) of entrapped small molecules, and has been the subject of experimental\textsuperscript{19, 22-25} and theoretical investigations\textsuperscript{26-30}. In general, drugs whose diameter is smaller than the average pore size can quickly diffuse through the structure of the hydrogel, but drugs whose diameter is equal to or greater than the average pore size, are drastically slowed down\textsuperscript{26, 31}. The movement of a drug molecule through a hydrogel’s non uniform pore structure is also related to the size of the “pore throat”\textsuperscript{25, 32}, that is the maximum size that a molecule can be and still travel in the hydrogel without limitations. In order to deliver drugs to a target site effectively, it is important to understand the relationship between the pore structure and drug size, the goal being to make hydrogel networks with appropriate pore size distributions and thus minimize premature drug loss.
Hydrogels prepared from poly(ethylene glycol) (PEG) have been used in a wide range of biomedical applications because PEG is a water soluble and non-toxic polymer. Among the various shapes of PEG hydrogels, those formed by multi-arm PEGs are useful when a highly homogeneous internal structure is desired. When the end of the PEG is modified with DNA and complementary DNA is added as a potential crosslinker, a hydrogel will form between these “precursors” as a result of the hybridization between DNA and its complement. Hydrogels formed in this way have the advantage that all the components are biocompatible, that the crosslinking process is spontaneous, and that the crosslinks can be reversibly disrupted thermally and by competitive interaction with other complementary DNA strands. In instances where DNA aptamers are used, crosslinking can also be reversibly disrupted by the presence of the DNA aptamer.

We are engaged in a computational and experimental research project which aims to develop DNA-enabled micro- and nano-sized hydrogels formed by multi-armed PEG molecules. This paper is the second in a three-part series. The first paper describes the development of a coarse-grained (CG) model for simulation of DNA hybridization. In this, the second paper, we simulate the formation of the DNA-mediated hydrogel using the developed model, and assess the hydrogel’s drug loading ability. In the third paper, we will simulate the degradation of the hydrogel when it encounters target molecules, thereby releasing previously-loaded drugs.

The goal of this research is to understand how the structure and concentration of the PEG precursors affect the formation and structure of hydrogels, and to predict the size of therapeutic molecules that can be entrapped and/or transported through them. The characteristics of the hydrogel structure (pore size) are determined by the precursor concentration, precursor structure, and degree of crosslinking. The sizes of representative drugs are known to be the following: the
hydrodynamic radius of rituximab is 54 Å, trastuzumab is 69 Å, ranibizumab is 28 Å, and aflibercept is 37 Å\(^45-46\). Rituximab is a chimeric monoclonal antibody used to treat several types of cancer and autoimmune disorders. Trastuzumab is a monoclonal antibody that binds the HER2 receptor which is used for the treatment of breast cancer. Ranibizumab is a monoclonal antibody fragment that acts as an inhibitor of angiogenesis for the treatment of wet age-related macular degeneration. Aflibercept is a recombinant protein that also acts as an angiogenesis inhibitor. It is used for the treatment of macular degeneration and metastatic colon cancer. Here, we investigate (1) how the precursor concentration required to form a hydrogel depends on the number of arms of the precursors, (2) how the pore diameter depends on the precursor concentration and number of arms, and (3) how the pore structure affects the movement of molecules through the hydrogel. The benefit of this work is that we can find design parameters for preparation of DNA-crosslinked hydrogels with desired pore size so as to optimize drug retention and controlled release.

In this project, we conduct discontinuous molecular dynamics (DMD) simulations to model the formation and drug carrying ability of oligonucleotide-crosslinked hydrogels. The PEG precursors are covalently functionalized on their ends with oligonucleotides. Two different structures of oligonucleotide-functionalized PEG precursors are modeled in CG representation: 4-armed and 6-armed. The crosslinker is a CG oligonucleotide that is complementary to the oligonucleotide on the end of the PEG precursors. The precursors are hybridized with crosslinker oligonucleotides to form a hydrogel network. Water is treated implicitly. The precursors are randomly located in the simulation box initially, and spontaneous network formation is observed as the simulation proceeds. The percolation probabilities of the networks formed at various precursor concentrations are analyzed to determine the lowest precursor concentration required
for hydrogel formation for each precursor shape and molecular weight. The pore size diameter distributions are calculated to learn how the shape, molecular weight, and concentration of precursors influence the hydrogel porosity. In addition, based on analysis of the pore size distribution, we find the maximum size molecule that can travel through the hydrogel networks without being trapped, i.e. the pore throat diameter. Finally, to understand the effect of pore size distribution on the migration of molecules in the network, the mean square displacements (MSD) of different-sized spheres in the percolated networks are investigated.

Highlights of the results include the following. The network structures crosslinked by different types of precursors are simulated using the CG representation. The network formed by the 4- and 6-armed precursors exhibits a high degree of crosslinking for all precursor concentrations investigated. The 6-armed PEG requires lower precursor concentrations to form a hydrogel than the 4-armed PEG regardless of the precursor molecular weight. The pore size distribution becomes narrower with higher PEG molecular weight, number of arms, and concentration, and the average pore diameter decreases accordingly. All the percolated networks have heterogeneous porous structures, which limit the movement of the molecules within them. Molecules with a diameter equal to or smaller than the pore throat diameter cannot be confined inside the hydrogel matrix. Lastly, the diffusion coefficients of inserted spheres with various diameters are calculated from the MSD data, and the reduction in movement for each size sphere is estimated. The MSD results for small spheres inserted in the percolated networks demonstrate the potential of oligonucleotide-mediated hydrogels for use as drug delivery vehicles.
3.2 Model and method

Each CG ethylene oxide (EO) repeat unit (CH₂CH₂O) of PEG is represented by a single sphere that contains 2 carbons, 4 hydrogens, and 1 oxygen. PEG molecules are modeled as linear chains of connected EOs. Two types of bonds are applied to maintain the connectivity and stiffness of CG PEG: covalent bonds and pseudobonds. The covalent bond is a real bond between adjacent CG EOs, and the pseudobond is an artificial bond between nonbonded CG EOs. Pseudobonds are imposed between a CG EO and the 2nd-nearest neighbor EO and between a CG EO and the 3rd-nearest neighbor EO. (The 2nd-nearest neighbor indicates EOs separated by one EO, and the 3rd-nearest neighbor indicates EOs separated by two EOs along the chain.) The distance between the CG EOs fluctuate between maximum and minimum bond length values. When the bond length reaches the minimum or maximum distance, an infinitely-repulsive force is exerted so that they return to the proper bond length range\(^ {47-48} \).

The values of the minimum and maximum distances between bonded CG EOs are determined by performing atomistic simulations. Simulations of united-atom PEG (Mw : 546.65 Da, number of repeat unit : 13) are performed with the GROMACS package to obtain the bond length distributions. Molecular topology information was obtained by using Automated Topology Builder (www.atb.com). The Lennard-Jones interaction parameters for the atoms in PEG are taken from Hezaveh’s research data\(^ {49} \). Fifteen PEG chains were placed in a box of 100 Å × 100 Å × 100 Å, and 32548 water molecules were added to fill the box. Ninety six sodium cations and ninety six chlorine anions were inserted to mimic 160 mM salt conditions. The simulation was conducted for 20 ns at 310 K and 1.0 bar (NPT ensemble). The various bond length distributions were obtained by collecting the center of mass distance between the CG EOs. CG distributions for the covalent bond lengths, and for the bond and torsional angle
pseudobond lengths similar to those in the atomistic simulation can be achieved by limiting the minimum and maximum distances to a proper range. The minimum real bond and pseudobond lengths were selected by finding the smallest distance at which the bond distribution function reaches 30% of its maximum peak value, and the maximum real bond and pseudobond lengths were determined by finding the distance (larger than the maximum peak distance) where the bond distribution function reaches 30% of its maximum peak value. The so-determined CG bond lengths are as follows. The bond length between covalently-bonded CG EOs fluctuates with a minimum length of 2.985 Å and a maximum length of 3.685 Å. The pseudobond length for the bond angle has a minimum of 4.945 Å and a maximum of 6.935 Å, and the pseudobond length for the torsional angle has a minimum of 7.505 Å and a maximum of 10.025 Å.

The radial distribution functions (RDF) between two intermolecular CG EOs are obtained from center of mass data and used to define interaction parameters. The interaction between CG EO molecules are represented using the hard-sphere potential; this is because the EO-EO radial distribution function has no noticeable local maximum as would have resulted if there were molecular attractions. The hard sphere potential, \( u_{ij} \), is

\[
  u_{ij}(r) = \begin{cases} 
  \infty & 0 < r \leq \sigma_{ij} \\
  0 & \sigma_{ij} < r 
  \end{cases}
\]

where \( r \) is the separation distance, and \( \sigma_{ij} \) is the sphere diameter between two EOs \( i \) and \( j \). The hard-sphere diameter represents the minimum possible distance between a pair of non-bonded CG sites, and is determined by finding the shortest separation distance as reflected in the RDF between the CG sites. The hard sphere diameter is determined to be 3.125 Å.

Each nucleotide is modeled as a single CG interaction site to represent sugar, phosphate, and base. The model has four different types of nucleotides (adenine, thymine, cytosine, and guanine). According to the Watson-Crick base pairing rule, only adenine-thymine and cytosine-
guanine interactions need to be taken into account. The interactions between CG nucleotides are represented using the square-well potential; the model parameters were established in our previous research.44

Figure 3.1 shows the structures of the hydrogel precursors and crosslinker oligonucleotides considered in the coarse grained representation. The hydrogel precursors contain two distinctive moieties, PEG (black) and oligonucleotide (blue and green). There are two different types of PEGs; 4-armed, and 6-armed. The number of EO beads in the branch is determined by the molecular weight of PEG. PEGs with molecular weights of 2, 5, and 10 kDa were selected. Table 1 shows the number of EO repeat units on one branch for the different molecular weights and shapes.

The oligonucleotides are attached to the end of the PEG arms and act as crosslinking sites. Two different oligonucleotide sequences are used: 5’ GGACGGTGCAGGCG 3’ (DNA (A), blue in Figure 3.1) and GTGACTGGACCCCC (DNA (B), green in Figure 3.1). These sequences are chosen because their melting temperatures are higher than body temperature. Thus, the given oligonucleotides could be used for crosslinking since the hybridized state between those oligonucleotides and their complements would be maintained in the human body. The PEG precursor functionalized by the oligonucleotide sequence GGACGGTGCAGGCG is called Precursor A, and the PEG precursor functionalized by GTGACTGGACCCCC is called Precursor B. The sequence of the crosslinker oligonucleotide is 5’ CCTGCCACGCTCCGCCCCACTGGGGG 3’, which is the concatenation of the complementary oligonucleotides of the two oligonucleotides described above (schematically represented as DNA (A’) and DNA (B’) in Figure 3.1). Base pairings between those
oligonucleotides are the driving force for crosslinking. The crosslinking is formed by connecting Precursors A and B via the crosslinker.

Discontinuous molecular dynamics (DMD) simulations is used to simulate the formation of hydrogels and the movement of spheres (model therapeutics) through the hydrogel. DMD is a fast alternative to traditional MD that is applicable to systems of molecules interacting via discontinuous potentials. Because the discontinuous potential forces on the particles are exerted only when the particles collide, the computational costs of DMD are lower than in traditional MD, and this allows the study of longer time scales and larger systems. The DMD algorithm calculates the collision times among all the pair-wise collisions and advances the time to the point that the soonest collision occurs. Unlike conventional MD, which uses a constant time step, the time step of DMD is different in every calculation, so that the progress of the simulation is expressed in terms of the number of collisions.

The details of the CG DMD simulations of the hydrogel formation are as follows. Initially, 20 Precursor A and 20 Precursor B chains are randomly placed in a box. The numbers of crosslinker oligonucleotides in the box are set to be 100 and 140, respectively, for 4-armed and 6-armed precursors (1.5 molar excess relative to DNA (A) and DNA (B)). The lengths of each side of the box were selected to mimic precursor concentrations of 0.066, 0.158, 0.532, 1.038, 1.550, 1.936, 2.461, and 3.195 mmol/L. The concentrations are calculated simply by $c = N_{\text{precursor}}/(V_{\text{box}} N_A)$, where $N_{\text{precursor}}$ is the number of precursors in the simulations, $V_{\text{box}}$ is the volume of the simulation boxes, and $N_A$ is Avogadro’s number which is needed for unit conversion. The box lengths were 1000, 750, 500, 400, 350, 325, 300, and 275 Å for the respective precursor concentrations. The temperature is maintained constant by using the Anderson thermostat\textsuperscript{50}. Simulations are performed at body temperature (310 K) because the
hydrogel’s drug transport should take place in the body. The simulation body temperature is taken from our previous paper \( (T_{\text{body}} = 0.58)^{14} \). In actual simulations, the temperature is cooled down by 0.1 for every 1 million collision from an initial very high temperature of \( T_s = 8.0 \) until this temperature was achieved. The purpose of using high temperature at the start of the simulation is to spread the precursors uniformly around the simulation box before the actual “reaction” occurs.

The percolation probability is used to determine if the network meets the minimum requirements to be a hydrogel or not. We defined an aggregate to be at least two precursors that are connected by a crosslinker. The aggregate is considered a “percolated network” if there is a connected path in the aggregate that spans from one end of the simulation box to the other along a path that connects to its own periodic image in one direction. Although percolation in one direction is a prerequisite for being a gel\(^{51} \), percolation in three dimensions seems to be necessary for this study in order to have a hydrogel with spaces for storing small molecules (Supplementary Figure 3.1a). A network percolated in the x and y but not the z directions would be a 2-dimensional slab (Supplementary Figure 3.1b), and a network percolated in one direction would be an infinitely long string polymer. We set the maximum percolation probability value to be 3, which adds up to the maximum percolation probabilities in the x, y, and z directions to describe the percolated state of network. The three-dimensional percolation probability, \( \Pi \), is defined as

\[
\Pi = \frac{\langle N_{\text{perc},x} \rangle + \langle N_{\text{perc},y} \rangle + \langle N_{\text{perc},z} \rangle}{\langle N_{\text{config}} \rangle}
\]

where \( N_{\text{perc},i} \) is the number of configurations which have a network that is percolated in \( i \) direction and \( N_{\text{config}} \) is the total number of configurations investigated. The angle brackets denote an average over time. The percolation probability is recorded every million collisions after the
system reaches the equilibrium state, and the percolation probability value at a given temperature is the average from the probabilities of 5 different simulations at that temperature.

The pore diameter distributions in the percolated network are calculated to quantify its structure. To do this, a random position is selected in the pore and the largest sphere that encompasses that random position is found. The diameter of that sphere is chosen as the diameter of the pore. The simulations for obtaining the pore diameter distributions are conducted for 50 million collisions starting with an already-equilibrated network as an initial configuration. Every 1 million collisions, 2000 random locations are used to measure the pore diameters.

To understand the effect of the pore structure on the migration of molecules in the network, a sphere is inserted into the percolated network and its mean square displacement (MSD) is calculated. The sphere has no interactions with the EOs or with the nucleotides; the point is just to focus on how network structure affects the MSD. The sizes of sphere inserted into the hydrogel network are 55, 60, 65, 70, and 75 Å. The molar mass of all the spheres is set to be 1.0 g/mol so that the MSD variations become a function of sphere size only. The MSD was determined from the average of five simulation runs. The starting positions of the spheres were the same for all the MSD calculations. The directions and magnitudes of the initial velocity of the sphere were different in every simulation, but the values were chosen so that the kinetic energy of the spheres corresponds to body temperature. The MSD is defined as

$$\text{MSD}(\Delta t) = \langle (r(t + \Delta t) - r(t))^2 \rangle$$

where \(\Delta t\) is the time interval and \(r(t)\) is the position of the sphere at time \(t\). The angle bracket denotes an average over time. The diffusion coefficient is obtained from the slope of the MSDs vs. time graph.
3.3 Results and discussion

Self-assembly of precursors in presence of crosslinker

Figures 3.2 a and b show snapshots of the initial and final configurations for the systems of 6-armed/10 kDa precursors. At the initial configuration, the precursors of A (blue), B (green), and crosslinker strands (yellow) are randomly placed in compact shapes. After 8 billion collisions, Precursor A and B molecules assemble together by hybridizing with the crosslinker strands. The red artificial bonds are used to illustrate the hybridization between precursors and crosslinker strands; chains of connected red bonds which look like ladders indicate complete hybridization. The precursors A and B are evenly distributed throughout the simulation box, and several void spaces which represent the pores of the hydrogel network are observed.

The rate of assembly of the system is quantified by measuring the fraction of precursors and crosslinkers that are hybridized. The hybridized fraction in the equilibrium state is defined as the number of hybridized nucleotide pairs between precursors and crosslinkers divided by the total number of possible nucleotide pairs in the system. Figure 3.2c shows the fraction of total hybridized pairs of 6-armed/2 kDa precursors at various concentrations from 0.066 to 3.195 mmol/L. At the beginning of the simulation, the hybridized fractions are low at every concentration. However, as the simulation proceeds, crosslinkers pair with their complementary oligonucleotide sequences on the precursors and the hybridized fraction increases. The hybridized fractions reach a plateau after about 4 billion collisions; the values for 4-armed and 6-armed precursors are 0.82 – 0.88 and 0.86 – 0.90, respectively, at all examined concentrations.

The degree of saturation is defined as the number of crosslinkers that are hybridized to the precursors divided by the number of crosslinkers needed to fully crosslink the network. Table 2 displays the degree of saturation for the different shaped-precursor systems in the equilibrium
state. The degrees of saturation for the 4-armed and 6-armed precursor systems are close to 90%. The 4-armed and 6-armed precursors achieve higher high levels of crosslinking regardless of the molecular weights.

**Three dimensional percolation; Formation of hydrogel**

To check whether the clusters that form in the simulations are truly interconnected networks or just disconnected aggregates, the three-dimensional percolation probabilities of the crosslinked structure are calculated as a function of precursor concentration. Recall that unlike the conventional percolation probability which ranges from 0 to 1, the maximum value of the percolation probability in our simulation is 3, which is the sum of the maximum percolation probabilities in x, y, and z directions. Having a value of 3 means that the system has percolated in each of the x, y, and z directions, and that there are void spaces surrounded by network skeleton. The value of percolation probability is determined by averaging the results of five simulations performed at the same concentration. Figures 3.3 a and b show the percolation probabilities of crosslinked structures formed by 4-armed and 6-armed precursors with molecular weights of 2, 5, and 10 kDa as a function of concentrations. Once again, the length of a single arm for each precursor is: for 4-armed precursor, 11 EOs at 2 kDa, 28 EOs at 5 kDa, and 57 EOs at 10 kDa; for 6-armed precursor, 8 EOs at 2 kDa, 19 EOs at 5 kDa, and 38 EOs at 10 kDa. Note that the heavier precursors have longer PEG lengths. Regardless of the number of arms, precursors with longer lengths percolate at lower concentrations. This implies that larger precursors are better at forming a percolated network than the shorter precursors at the same concentration. The minimum precursor concentration required to form a three-dimensional percolated network is ~3.0 mmol/L for 4-armed precursors. In the case of the 6-armed
precursors, three dimensional percolation is first observed at a concentration less than 1.0 mmol/L, relatively low compared to the 4-armed precursors. A noteworthy point is that the degree of percolation for 6-armed precursors increases abruptly at a certain concentration, rather than gently as in networks formed by the 4-armed precursor. In summary, the higher the number of branches in the precursor and the longer their length, the lower the concentration required to form a three-dimensionally percolated structure.

**Pore diameter distribution**

The percolated networks contain empty spaces, pores, surrounded by crosslinked precursors; the size of these pores depends on the precursor concentration, molecular weight, and structure. High precursor concentration makes the hydrogels more likely to have relatively uniform small-sized pores. Figure 3.4a displays the pore diameter distributions for percolated networks of 6-armed /10 kDa precursor at different precursor concentrations. The systems are at higher precursor concentrations than the three dimensional percolation threshold. In all graphs of the pore diameter distribution, very small pores (< 10.0 Å) are commonly found with a high probability. These are formed when several chains cluster together to form bundles, creating innumerable small voids and correspond to the first peak in the figure. A better measure of the topology of the available pore space is the diameter associated with second peak in Figure 3.4, the local maximum occurring after 10 Å. We will refer to this as the “characteristic diameter”. The three-dimensionally percolated networks formed at relatively low precursor concentrations have a broad pore diameter distribution profile. As the precursor concentration increases, the distribution profile gradually narrows and peaks at a higher probability. The characteristic pore diameter decreases as the concentration increases, and ranges from 55 Å at 1.04 mmol/L to 17 Å
at 3.20 mmol/L. Supplementary Figure 3.2 shows the pore distributions at a variety of precursor concentrations for all types of precursors. In all cases, the shapes of the pore diameter distributions are narrow and the characteristic pore diameters are small at high precursor concentrations.

The molecular weight of the PEGs also affects the pore structure of the percolated networks. Figure 3.4b displays the effect that increasing the precursor’s molecular weight has on the pore distribution at a fixed precursor concentration (3.20 mmol/L) and shape (6-armed). The distributions become narrower as the molecular weight increases with characteristic pore diameters of 26, 21, and 17 Å for 2, 5, and 10 kDa molecular weights respectively. Similar results are seen for the rest of the pore distribution curves of percolated network generated by 4-armed and 6-armed precursors (Supplementary Figure 3.2). Because the volume of the system is constant, the larger the molecular weight of the precursor, the smaller the volume occupied by the pores, and eventually the pore size decreases. Thus, a negative correlation between the size of the pores and molecular weight of the precursor is observed.

Lastly, precursors with many branches are found to be advantageous for forming uniform porous structures. The influence of the number of branches on the pore size distribution can be seen by comparing the pore diameter distributions for the 4- and 6-armed precursors at fixed precursor molecular weight and concentration. Figure 3.4c shows the pore distributions of the 4-armed /10 kDa (black) and 6-armed /10 kDa (red) networks at precursor concentration of 2.46 mmol/L. At a fixed concentration and molecular weight, the network formed by 6-armed precursor has a slightly narrower pore diameter distribution than the network formed by the 4-armed precursors. To summarize the three points, relatively narrow pore diameters are
established in the three-dimensionally percolated network under the conditions of high precursor concentration, large molecular weight and more branches.

**The non-uniformity of the pore structure**

The non-uniformity of the pore structure limits the movement of molecules located in the percolated networks. Obviously since there is a distribution of pore diameters, there will be pores that molecules of a certain size can reach and pores that they can’t reach. Pores of a certain size that are reachable within the hydrogel can be identified by placing virtual spheres of that size in many random locations of the simulation box without overlapping with the hydrogel skeleton. The shape of a pore within the hydrogel can be visually expressed as a conglomeration of the virtual pores lumped together. The dark blue areas in Figure 3.5b show the pore structure that small molecules of 40.0 Å diameter can travel through, and those in Figure 3.5c show the pore structure that large molecules of 60.0 Å diameter can travel through. Comparing the two figures, the volume that molecules of 40.0 Å diameter can travel through is much larger than the volume that molecules of 60.0 Å diameter can travel through. In addition, the space in which small molecules of 40.0 Å can move is connected to its own periodic images in the x, y, and z directions; i.e. the pore structure is in a percolated state. On the other hand, the space in the hydrogel where large molecules of 60.0 Å can move is disconnected. Therefore, relatively small molecules in the hydrogel are allowed to move almost everywhere in the gel matrix, while relatively large molecules have very limited movement because of the hydrogel’s non-uniform disconnected pore structure. It follows that there exists a maximum size molecule that can move through the inside of the hydrogel. This is the “throat diameter” of the pore.
The existence of the pore throat

The existence of the pore throat diameter is confirmed by observing cross-sectional images of the percolated network. Figure 3.6 shows the y-z cross-section of percolated network formed by 6-armed /10 kDa with precursor concentration of 1.55 mmol/L at x = -50.0 Å. Different intensities of blue represent pores in which molecules of different diameter from 20.0 to 60.0 Å can travel, respectively. The pore area where a molecule with a diameter of 20.0 Å can travel is distributed throughout the network. As the molecule size increases, the area accessible to molecules of that size decreases gradually; very limited areas are allowed for the motion of molecules with diameters larger than 60.0 Å. Thus, the throat diameter of a pore is determined by investigating the pore connectivity of different diameters.

The pore throat diameter of the hydrogels

The throat diameter of hydrogels decreases as the precursor concentration increases. To determine the throat diameter, virtual beads of 10 different diameters (from 35 to 80 Å in 5 Å intervals) are placed in the pore region of percolated networks. We observe whether the virtual beads are connected in the x, y, and z directions. The connectivity of the virtual beads is verified using the percolation probability concept. The percolation of the virtual beads is quantified in terms of a pore connectivity probability for better understanding. The pore connectivity probability is defined to have a value of 1 if the virtual beads form a chain that percolates across the simulations box regardless of the direction, and 0 otherwise. Figure 3.7a is a graph comparing the pore size distribution (black) and the pore connectivity probability (blue) of a hydrogel prepared using 6-armed/10 kDa at a precursor concentration of 1.55 mmol/L. The characteristic pore size for this case is 39 Å. The pore connectivity probability is constant at 1 up
to a pore diameter of 55.90 Å, which means that the pores are percolated. However, the pore connectivity probability drops to zero at 60.20 Å, which means that the pores lose connectivity. From this, we determine that the throat diameter is 55.90 Å which is the largest diameter that maintains pore percolation. The pore throat diameter of the hydrogel can be adjusted to be 16 to 170 Å by varying the molecular weight, the number of branches, and the concentration of the precursors (Supplementary Figure 3.3). Figure 3.7b is a graph comparing the characteristic pore diameter and the pore throat diameter for the hydrogel made using the 6-armed/10 kDa precursor at various concentrations. The characteristic pore diameters and the pore throat diameters of the percolated networks decrease as the precursor concentration increases. When the precursor concentration is less than 2.46 mmol/L, the throat diameters are greater than the characteristic pore sizes. However, when the precursor concentration is above 2.46 mmol/L, the throat diameter becomes similar to the characteristic pore diameter suggesting that the connected pores of the hydrogel are like pipes with a constant diameter. In other systems, except 6-armed/10 kDa, the pore throat diameter is always larger than the characteristic diameter at the precursor concentrations considered. Molecules placed inside the hydrogel whose characteristic size is greater than or equal to the throat diameter of the pore will be completely trapped in the hydrogel matrix. Ultimately, to deliver the drug to its destination without loss, a hydrogel with a smaller throat diameter than the size of the drug should be used.

**The diffusion of molecules within the network**

The movement of small molecules in the percolated networks (hydrogel) is more limited than those in the unpercolated networks when the molecular weight and number of precursor arms are fixed. To investigate the movement of a molecule inside the network, hard spheres of different sizes are inserted in the percolated network and the MSD of the spheres are measured.
For the measurement of MSD, two networks with 6-armed/10 kDa at 0.16 and 1.55 mmol/L are selected; one represents the unpercolated state and the other represents the percolated state. The diameters of the spheres are 50, 55, 60, 65, 70, and 75 Å which are the size corresponding to the diameter range of ranibizumab (56 Å) and aflibercept (76 Å). Note that the mass of the sphere is constant as 1 g/mol regardless of the diameter in order to maximize the movement of sphere per unit time and avoid mass effects in the MSD. The MSDs of the spheres are measured by inserting them one by one into the network and tracking their positions. Figure 3.8 shows the results of the MSD measurement of different sized spheres in the two networks. The MSDs of spheres in the unpercolated network are nearly the same for the two sphere diameters (Figure 3.8a). This is to be expected because the unpercolated network does not have pores surrounded by precursors. However, in the percolated network, the MSDs are significantly reduced compared to those in the unpercolated network regardless of the sphere diameter (Figure 3.8b). This implies that once three dimensional percolation is established, the movement of the molecule inside the network is constrained by collisions due to the structure of the network. The MSDs of relatively small-sized spheres are high even though the sphere is in the percolated network because they are small enough to move freely without being disturbed. As the sphere diameter increases, the MSDs decrease because the available void space becomes too small.

The movement of spheres in the three-dimensionally percolated network is quantified by calculating the diffusion coefficient from MSD data. Table 3 shows the calculated diffusion coefficients of all size hard spheres measured in the unpercolated and percolated networks. The diffusion coefficients of the spheres in the unpercolated networks are $3.78 \times 10^{-1} \text{Å}^2/\text{reduced time (t*) or more}$; the diffusion coefficients of the spheres in the percolated networks are lower. Diffusion coefficients calculated at all concentrations are inversely correlated with the diameter
of inserted spheres. The diffusivities of spheres with diameter 55 Å, 65 Å, and 75 Å in the percolated network (6-armed/10 kDa, at 1.55 mmol/L) are about 2.30 %, 0.84 %, and 0.09 % of the diffusivity of spheres in the unpercolated network.

3.4 Conclusion

We describe the results of discontinuous molecular dynamics simulations of the formation and the structural properties of an oligonucleotide-crosslinked PEG based hydrogel. The hydrogel networks are formed by 4- and 6-armed PEG precursors that are covalently functionalized on all ends with oligonucleotides. The crosslinker is a single oligonucleotide whose sequence is complementary to the oligonucleotide conjugated to the precursors. Network formation is achieved by hybridization between the precursor’s oligonucleotide moiety and the crosslinker. The formation of a network was investigated in the concentration range from 0.066 to 3.195 mmol/L for each precursor. High levels of crosslinking are achieved for the 4- and 6-armed precursor systems.

When the network is percolated in the x, y, and z directions at a given concentration, it is believed to satisfy the minimum requirements for being a hydrogel. We calculated percolation in three dimensions because it ensures the formation of a hydrogel with cargo space for small molecules. The simulation results show that the likelihood of three-dimensional percolation increases as the precursor concentration increases for all shape precursors. The concentrations needed to form a stable percolated structure are 3.0 and 1.0 mmol / L for the 4- and 6-armed precursors, respectively, regardless of the precursor molecular weights. As the number of branches of the precursor increases, lower precursor concentrations are sufficient to achieve three-dimensional percolation. We can understand this by imagining how an ideal hybridized
structure for each shape precursor would look. When they are crosslinked ideally, 4- and 6-armed precursors can make a diamond lattice and a face-centered cubic lattice, respectively, so those precursors are able to create a three-dimensional structure intrinsically. However, as the 6-armed precursor has more hybridization sites than 4-armed precursor, crosslinking can be achieved more easily. Thus, 6-armed precursors form stable hydrogel networks at low concentrations.

Next, the pore diameter distributions within the percolated network structures formed by 4- and 6-armed precursors were analyzed. The pore diameter distribution depends on the shape, molecular weight, and concentration of the precursors. In order to form a hydrogel with narrow pore size distributions, a precursor should have many branches and high molecular weight and should be crosslinked at a high concentration. As the pore sizes in DNA-mediated hydrogel are not uniform, the maximum size of a material that could travel freely within the hydrogel should be determined. This maximum size is called the pore throat diameter. The pore throat diameter of the hydrogel is found to be higher than the characteristic diameter (the most frequently observed pore diameter > 10 Å) for most of the investigated structures and precursor concentrations. The reason why the throat diameter is larger than the characteristic diameter is because the small-sized disconnected pores (or separated chambers), which are not associated with connected pores, are included in the pore diameter distribution curve. The hydrogels crosslinked by 6-armed/10 kDa precursor at a concentration of 2.46 mmol/L or higher have a pore throat diameter that is similar to the characteristic diameter. The reason that those two diameters are similar at high concentrations is that the sizes of the voids within the pores have become more uniform. Measuring the distribution of connected and disconnected pores would be a new research topic for a deeper understanding of the structure of hydrogels. When the hydrogel is used as a drug
delivery vehicle for loading molecules that are smaller than pore throat diameter, immersion of the hydrogel in a liquid to load drug molecules into the hydrogel by equilibrium partitioning would work. However, at the same time, it would be possible for the molecules to escape from the hydrogel before it reaches the target site.

Lastly, the drug carrying ability of the percolated network (hydrogel) was verified by analyzing the diffusion of spheres in the networks. Spheres of various size diameters (55, 60, 65, 70, and 75 Å) were placed in percolated networks formed by 6-armed precursors at 1.55 mmol/L precursor concentration, and the MSD of the beads was calculated. As a control, the MSD of the same sized beads in a network formed at 0.16 mmol/L, an unpercolated network, was measured. The MSDs of the various sized beads in the unpercolated network with 0.16 mmol/L are similar because the void space is larger than the size of the inserted beads. On the other hand, the MSDs of all size beads in the percolated network (1.55 mmol/L) decrease because the movement of beads is reduced by colliding with the scaffolds of the networks. As the bead diameter increases, the MSD gradually decreases because the pore space becomes too small for the beads to move in the hydrogel network. When the size of the bead is greater than the pore throat diameter, the MSD is relatively independent of the bead size, indicating that the material is entrapped. For example, ranibizumab (a drug for macular degeneration with hydrodynamic diameter: 55.2 Å) can be carried by the hydrogel formed with 6-armed/10 kDa at the concentration of 1.94 mmol/L since its pore throat diameter is 43.0 Å.

The main conclusions that can be drawn from this simulation study of oligonucleotide-crosslinked hydrogels are the following: 1) The required concentration for the formation of the hydrogels by various precursor shapes is predicted to be; 3.0 mmol/L for 4-armed precursors and 1.0 mmol/L for 6-armed precursors. 2) The structure of the formed hydrogel can be understood
through the pore diameter distribution and pore connectivity analysis; pore size distributions indicate the extent to which the pore structures are non-uniform and the pore connectivity probability allows determination of the pore throat diameter. 3) The drug-carrying ability of the hydrogel can be analyzed by measuring the MSD of small molecules in the hydrogel; percolated networks show lower MSDs than unpercolated networks. 4) The size of drugs should be larger than the pore throat diameter to deliver them without leaking. There results could be used to design DNA crosslinked hydrogels for drug delivery application by adjusting the structure and concentration of the precursors to control porosities.

In next stage of this project, we will develop a model of a target molecule (adenosine) that triggers the DNA crosslinks to unhybridize, and consequently the hydrogel to degrade and release drugs, by inclusion of a specific oligonucleotide (aptamer) sequence as part of the crosslinker. The crosslinker will be extended to include an aptamer sequence to react with the target. Assuming a situation where the hydrogel meets the target molecule, the adenosine will be introduced into the hydrogel network. The hydrogel network generated by the simulation will be used as an initial configuration to observe interaction with molecular targets.

Although our CG simulations provide molecular-level understanding of the formation and drug carrying ability of oligonucleotide-crosslinked hydrogels, the model has several limitations. First, the drug molecule is designed not to have any interactions with the hydrogel components. In reality, drug molecules would interact with the PEG or the nucleotide via electrostatic interactions, hydrophobic interactions and/or hydrogen bonding, affecting the movement of drug molecules. Second, water is treated implicitly in our simulation and the absence of hydrodynamics would have distorted the structure of the hydrogel somewhat differently from the actual one. Finally, in order to predict conditions under which a hydrogel
would actually form, we need to analyze not only the network percolation but also its mechanical properties such as the modulus.

3.5 Acknowledgement

This work was supported by the National Institutes of Health (EB006006), the NSF’s Research Triangle MRSEC on Programmable Soft Matter, DMR-1121107 and CBET 1512059, and the NSF Partnership for Research and Education in Materials (PREM), DMR-1205670.
3.6 References


Table 3.1. Number of CG EOs in one branch for each shape and molecular weight PEG precursor.

<table>
<thead>
<tr>
<th></th>
<th>2 kDa</th>
<th>5 kDa</th>
<th>10 kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-armed PEG</td>
<td>11</td>
<td>28</td>
<td>57</td>
</tr>
<tr>
<td>6-armed PEG</td>
<td>8</td>
<td>19</td>
<td>38</td>
</tr>
</tbody>
</table>
Table 3.2. The saturation degree of precursors with different shapes and molecular weights. Unit : %

<table>
<thead>
<tr>
<th></th>
<th>4-armed</th>
<th>6-armed</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 kDa</td>
<td>89.38</td>
<td>89.19</td>
</tr>
<tr>
<td>5 kDa</td>
<td>88.12</td>
<td>89.81</td>
</tr>
<tr>
<td>10 kDa</td>
<td>86.38</td>
<td>89.92</td>
</tr>
</tbody>
</table>
Table 3.3. Diffusivity of spheres with various diameters in unpercolated (0.16 mmol/L) and three dimensionally percolated (1.55 mmol/L) networks.

<table>
<thead>
<tr>
<th>Sphere size (Å)</th>
<th>Diffusivity (cm² / t*) in 0.16 mmol/L (unpercolated)</th>
<th>Diffusivity (cm² / t*) in 1.55 mmol/L (percolated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55.0</td>
<td>4.83 × 10⁻¹</td>
<td>8.69 ± 3.33 × 10⁻³</td>
</tr>
<tr>
<td>60.0</td>
<td>-</td>
<td>6.03 ± 0.84 × 10⁻³</td>
</tr>
<tr>
<td>65.0</td>
<td>-</td>
<td>3.18 ± 1.27 × 10⁻³</td>
</tr>
<tr>
<td>70.0</td>
<td>-</td>
<td>2.61 ± 1.25 × 10⁻³</td>
</tr>
<tr>
<td>75.0</td>
<td>3.78 × 10⁻¹</td>
<td>1.05 ± 0.56 × 10⁻³</td>
</tr>
</tbody>
</table>
Figure 3.1. Schematic illustrations of (a) 4-armed, (b) 6-armed precursors, and (c) a crosslinker.
Figure 3.2. The (a) initial and (b) final snapshots from a simulation containing 6-armed precursors (Blue – Precursor A, green – Precursor B) and crosslinker oligonucleotides (yellow). The inset figure of (b) is an enlarged representation of the hybridization of oligonucleotides; the base-pairing between the CG nucleotide beads is indicated by the red artificial bonds so that the hybridized state can be easily recognized. (c) The hybridized fractions of all oligonucleotides in precursors versus time. Curves of different colors indicate results obtained from different precursor concentrations.
Figure 3.3. Percolation probability in three dimensions versus the precursor concentrations for (a) 4-armed and (b) 6-armed precursor systems. Black, red, and blue curves in each figure represent precursors with molecular weight of 2, 5, and 10 kDa.
Figure 3.4. Pore diameter distribution for percolated network of (a) 6-armed/10 kDa at a variety of concentrations, (b) 6-armed/3.20 mmol/L at different molecular weights, and (c) 10 kDa / 2.46 mmol/L with different precursor structures.
Figure 3.5. (a) The shape of hydrogel formed by 6-armed/10 kDa at 1.55 mmol/L (blue – Precursor A, green – Precursor B, yellow – crosslinker, and red – artificial bond meaning hybridization). (b) and (c) show the positions of the pores (dark blue) in which molecules of 40.0 and 60.0 Å can move within the hydrogel (pink), respectively.
Figure 3.6. A contour map of pore of hydrogel formed by 6-armed/10 kDa at 1.55 mmol/L at x = -55.0 Å. The areas of the pores that can be reached by different sizes of molecules (20, 40, and 60 Å) are shown in blue colors with different intensities.
Figure 3.7. (a) Comparison of pore distribution and connectivity for percolated network by 6-armed/10 kDa at a precursor concentration of 1.55 mmol/L and (b) the characteristic pore diameters and pore connectivity thresholds of percolated network by 6-armed/10 kDa at various precursor concentrations.
Figure 3.8. Mean square displacement of spheres in (a) unpercolated network and (b) three dimensionally percolated networks. Each curve represents MSD of molecules of different diameters.
Figure S3.1. The projection views of (a) 3-dimensionally and (b) 2-dimensionally percolated networks with periodic duplications. Top to bottom x-y, y-z, and z-x planes. The blue box represents the simulation box.
Figure S3.2. Pore diameter distributions of networks formed by (a) 4-armed/2 kDa, (b) 4-armed/5 kDa, (c) 4-armed/10 kDa, (d) 6-armed/2 kDa, (e) 6-armed/5 kDa, and (f) 6-armed/10 kDa precursors.
Figure S3.3 Pore connectivity threshold of percolated network by different precursors.
Chapter 4

Development of a Coarse-Grained model of DSPE using multistate iterative Boltzmann inversion for discontinuous molecular dynamics simulations: an improved version of LIME

4.1 Introduction

Lipid bilayers result from the self-assembly of two phospholipid leaflets\(^1\). When dispersed in water, the hydrophobic moieties on lipid molecules aggregate to minimize their contact with the surrounding aqueous medium\(^{1-2}\). The lipid molecules spontaneously form a lipid bilayer, joining together to hide their hydrophobic tails from the surrounding water and exposing their hydrophilic heads. The structure of lipid bilayer is stabilized by the intermolecular van der Waals interactions\(^{3-4}\). Lipid bilayers are a common component of cell membranes and are essential for the life of living organisms\(^{5-6}\). They play an important role in many biological systems such as protein functions and cell signaling\(^{7-11}\). Lipid bilayers can form closed spherical membrane structures called liposomes\(^{12-13}\). Liposomes have components and structures that are very similar to animal and plant cells\(^{14-15}\), so they are used in studies of cell membrane function\(^{16-17}\). Because it is easy to add various functional groups to liposomes by modifying the constituent lipid molecules, liposomes are used in a variety of applications\(^{18-20}\).

Liposomes have recently attracted attention as drug delivery vehicles for use in treating a variety of diseases, especially cancer\(^{21-23}\). Many drugs, including DNA-based drugs, are not able to enter the target cell when administered directly into the bloodstream. An alternative is to internally store or electrostatically bind drugs to liposomes and then transport them directly to cells\(^{24-26}\). Hydrophilic drugs can be encapsulated in the aqueous phase of the liposome core and hydrophobic materials can be entrapped within the hydrophobic bilayer because of the lipids’
amphiphilic property. Fetterly et al. recently developed liposomes that incorporate paclitaxel, a well-known drug for cancer treatment. Similarly, Felgner and colleagues first reported that cationic liposomes would be useful for gene transfer, and they have studied gene therapy using liposomes since 1987.

Various simulation methods have been developed to explain the behavior of lipid bilayers or liposomes. These range from atomistic (high-resolution) simulations to coarse-grained (low-resolution) simulations. Atomistic simulations of lipid bilayers have been conducted to study the structure of lipid bilayers, the permeation of ions or molecules through a lipid membrane, and the interaction between materials and lipid bilayers. One limitation of atomistic simulations is that their detailed description of a molecular geometry and energetics makes them computationally intensive. Simulation efficiency can be improved by coarse-grained (CG) modeling in which groups of several atoms are combined together into a single interaction unit. Coarse-grained models are attractive for studying lipid bilayers because they can be used to examine large systems of molecules at very long time scales. The most widely used coarse-grained model is the Martini model developed by Marrink. The Martini model is able to reproduce lipid bilayer structure, dynamic properties, thermodynamic properties, and so on in a short time frame.

Another coarse-grained model for lipids is “LIME (Lipid Intermediate Resolution Model)” designed for use with discontinuous molecular dynamics (DMD) simulations. LIME is an implicit solvent model for 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) developed by Curtis and Hall, in which the intermolecular potential is a discontinuous function of the separation between two molecules. The model parameters in LIME were obtained by using a multiscale modeling procedure: radial distribution functions (RDF) between bonded and
nonbonded pairs of CG sites along the phospholipid chain taken from atomistic simulations are used to extract the LIME geometrical and the energy simulation parameters. The extracted energy parameters are “potentials of mean force”, which are the effective potentials (van der Waals interactions plus electrostatic interactions) between the groups of sites in water. DMD simulations using the LIME model, formed lipid bilayers in short computation times. LIME predicted physical properties of lipid bilayer that were comparable to those predicted by the atomistic simulations.

LIME had a few limitations: 1) the RDFs of the lipid bilayer calculated from the atomistic simulation do not match the RDFs from the coarse-grained simulation in detail, 2) LIME did not describe the intermolecular attractions in detail it used only a single square well potential to describe the interactions between coarse grained sites, and 3) the set of coarse-grained parameters extracted from atomistic simulations at one temperature are not necessarily accurate at other temperatures.

In order to overcome these limitations and to develop a more accurate CG lipid model for LIME, we introduce multiple square well intermolecular potentials and use Iterative Boltzmann Inversion (IBI) to find the interaction parameters. Models in which multiple square wells are used in conjunction with DMD simulations are the following. Rutkowski et al. analyzed the phase behavior of dipolar colloidal rods by expressing the potential between the colloidal particles in terms of three square wells\(^3^8\). Benner et al. developed a CG chitosan model using multiple square wells; the reproduced CG RDFs matched the RDFs from atomistic simulations almost exactly\(^3^9\). IBI derives coarse-grained potentials by optimizing a potential to match target RDFs from an atomistic simulation\(^4^0\). Recently, Moore et al. developed an extension to the IBI method to include target RDFs from multiple states, adding constraints to the potential
optimization process. By adding these constraints, they succeeded in extracting a single set of universal interaction parameters applicable to multiple temperatures in their study.

In this paper, we develop a more accurate set of LIME parameters by applying multiple square well intermolecular potentials and applying multi state IBI method to calculate the CG potentials. The improved LIME interaction parameters allow us to conduct CG-DMD simulation of lipid bilayers for drug delivery applications at a variety of temperatures. The model lipid for this study is 1,2-distearoyl-sn-glycero-3-phosphoethanolamine (DSPE). Unlike the previously derived LIME force field, the improved LIME uses multiple square wells to more faithfully represent the shape of RDFs in atomistic simulations. In addition, multi-state Boltzmann inversion is used to determine the interaction energies between CG sites, thereby expanding the ability of LIME to model lipid bilayers at any temperature. We show that the discontinuous molecular dynamics simulations using the improved LIME force field accurately represents the structure of a DSPE lipid bilayer over a given temperature range.

The highlights of our results are the followings: Different numbers of square wells are needed for each of the 6 distinct CG type on DSPE. Each intermolecular potential is expressed using 3 or 4 square wells depending on the shape of the atomistic RDFs. First, the interaction energies for a single state are found using the modified IBI scheme for discontinuous potentials at 298 K. The values of energies are obtained after 40 iterations. Then the IBI method is extended to conduct multi-state optimization. The intermolecular energies are adjusted to achieve simultaneous convergence of CG RDFs from multiple states to RDF from atomistic simulations. CG simulations of DSPE lipid bilayer using the potentials from multi state IBI are conducted at 298, 340, 370 K, and the characteristics of the lipid bilayer are well predicted at each
temperature. The bilayer thicknesses of CG lipid bilayer at various temperatures are measured, and they are comparable with that of atomistic lipid bilayer.

4.2 Model and method

The representation of a lipid molecule for the improved LIME model is the same as that used in the original LIME model. The improved LIME model has six coarse grained types to represent the 190 atoms that compose a DSPE molecule. Figure 4.1 shows the all-atom, united-atom, and coarse grained representations of a DSPE molecule. The coarse graining scheme massively reduces the number of sites, so that the DSPE molecule in improved LIME has 16 coarse-grained types. The coarse-grained DSPE is composed of a polar head group which includes an ethanolamine (blue), phosphate group (yellow), two ester linkages (red and orange), nonpolar hydrophobic alkyl group (cyan), and a terminal tail group (green). The mass of each coarse-grained site is the sum of the atoms’ molar mass within each coarse grained site. Each CG site has a unique interaction potential that is different than those of the other sites.

The CG parameters for the improved LIME model are extracted from the results of explicit-solvent NPT ensemble united-atom simulation. The simulation contains 128 DSPE lipids and 5888 water molecules. We used Gromacs simulation package (version4.5.4) with the GROMOS96 53a6 force field. The Lennard-Jones interaction parameters for DSPE lipid are taken from the simulation results of Qin et al.. The initial configuration of the system was a pre-formed bilayer. The Berendsen thermostat was used to maintain the system temperature throughout the simulation with a time constant of 0.1 ps. Three different temperatures (298, 340, and 370 K) were selected, which correspond the solid phase, solid-liquid phase transition temperature, and liquid phase. The simulations were run for 100 ns with a time step of 0.001 ps.
The pressure was kept at 1.0 bar. The trajectories of all atoms were collected for the final 20 ns with a time step of 0.001 ps and used to calculate the centers of mass for each CG site.

To maintain the connectivity and stiffness of CG DSPE, three types of bonds are used; covalent bonds, pseudobonds to maintain bond angle, and pseudobonds to maintain torsional angle. The covalent bond indicates a real bond between adjacent DSPE CG sites. The bond angle is maintained via a pseudobond, an invisible bond between a CG site and its next nearest neighbor CG site along the DSPE chain. The torsional angle is maintained via a pseudobond between a CG and its second nearest neighbor CG site. The lengths of all real bonds and pseudobonds fluctuate maximum and minimum values. An infinite repulsion force is exerted so that they return to the bond length when the bond length reaches the minimum or maximum distances.

A total of 41 real bonds and pseudobonds between CG sites along of DSPE are defined. The minimum and maximum bond lengths for those bonds are calculated from atomistic simulation results for the center of mass distance between bonded CG sites. In CG-DMD simulations, distributions for covalent bonds and pseudobonds similar to those in atomistic simulations can be achieved by limiting the minimum and maximum distance to an appropriate range. The minimum bond and pseudobond lengths were selected by finding the smallest distance at which the bond distribution function reaches 30% of its maximum peak value. The maximum pseudobond length was determined by finding the smallest distance (larger than the maximum peak distance) where the bond distribution function reaches 30% of its maximum peak value. Since our goal was to obtain one parameter set of bond lengths that is applicable to various temperatures, we calculated the bond distribution at three different temperatures (298, 340, and 370 K) and selected the minimum and maximum bond lengths of the bonded sites to
span the largest range among these three values. The determined minimum and maximum bond lengths for all bonded CG sites are in Table S.1.

The intermolecular interactions between CG sites of DSPE are represented using hard spheres and multiple square wells (or shoulders). The hard sphere diameter, $\sigma$, and square-wells widths, $\lambda$, are determined from the average of radial distribution functions between pairs of nonbonded coarse-grained sites obtained at three different temperatures in the atomistic simulations, $g_{A,\text{ave}}(r)$, where $A$ and ave stand for atomistic simulation and average value, respectively. Hard sphere diameters for each pair of interaction sites were chosen by locating the minimum nonzero separation distance between the two sites. The number of square wells or shoulders for each CG pair is set so that the RDFs of the CG simulations can reproduce the RDFs of the atomistic simulations; this number depends on the shape of the RDFs. In general, the first well width is selected to cover the range from the hard-sphere diameter to the smallest value of $r$ at which $g_{A,\text{ave}}(r)$ is approximately 50% of the maximum height of the first peak. The second well width is selected to cover the range from the end of the first well to the smallest distance (larger than the maximum peak distance) at which $g_{A,\text{ave}}(r)$ is approximately 50% of the maximum value of the first peak. If $g_{A,\text{ave}}(r)$ shows a distinct second peak, the third well width is selected the location at which $g_{A,\text{ave}}(r)$ is at the middle point between the first local minimum and the maximum of the second peak. The last well is selected to cover the range from the end of the third well to the cutoff distance. If $g_{A,\text{ave}}(r)$ does not a second peak, then the third well is chosen to cover the entire region out to the cutoff distance. Each intermolecular pair of CG site has 3 or 4 discontinuous square well potentials. If the atomistic RDFs of intermolecular interaction does not have noticeable local maximum, the interaction between the CG pair is represented using the hard-sphere potential.
The discontinuous interaction energies, \( \varepsilon \), between CG sites are determined by an iterative Boltzmann inversion scheme modified for discontinuous potentials. In the IBI method for a single temperature, the potential between the CG sites is chosen such that the RDFs of the CG simulations match the RDFs of the atomistic simulation. The initial approximations for all CG discontinuous potentials are zero. Those discontinuous potentials are updated according to
\[
U_{CG}^{(i+1)}(r) = U_{CG}^{(i)}(r) - k_B T \ln \frac{g_{CG}^{(i)}(r)}{g_A(r)},
\]
where \( U_{CG}(r) \) is the coarse-grained potential, \( i \) represents the current iteration, \( k_B \) is the Boltzmann constant, \( T \) is the absolute temperature, \( r \) is the separation between particles, \( g_{CG}(r) \) is the CG RDF, and \( g_A(r) \) is the RDF between pairs of nonbonded coarse-grained sites in the atomistic simulations. The IBI method for multiple temperatures is similar to that for a single temperature, but it has to be modified to cover \( N \) states. The initial potentials between CG types are zero. The potentials are updated according to
\[
U_{CG}^{(i+1)}(r) = U_{CG}^{(i)}(r) - \frac{1}{N} \sum_{j=1}^{N} k_B T_j \ln \frac{g_{CG,j}^{(i)}(r)}{g_A(r)},
\]
where \( j \) represents the state, \( g_{CG,j}(r) \) is the CG RDF for \( j \)th state, and \( g_A(r) \) is the atomistic RDF in the \( j \)th state\(^{41}\).

The details of the CG simulations used for the IBI procedure are as follows. Each CG simulation has the same box size as the atomistic simulations. A pre-formed lipid bilayer consisting of 128 DSPEs is used as the initial configuration in the CG simulations. The Anderson thermostat was used to maintain the temperature constant; in this method, the velocity of a CG bead is adjusted to keep the system’s Maxwell–Boltzmann velocity distribution consistent with the set temperature. The potentials are updated according to the Eq. (2) every 100 million collisions.
4.3 Results and discussion

Single state IBI (SS-IBI) is performed to find the values of square well depths between each coarse-grained type for DSPE at T=298 K. The total number of square wells between the 6 types of coarse grained sites is 45, and the iteration process is conducted at 298 K. The convergence percentage is defined as the number of square wells that converges the results of CG RDFs and atomistic RDFs over the total 45 square wells. Figure 4.3a shows the convergence percentage over the number of iterations. After 40 iterations, the convergence percentage reaches 100%. Figure 4.3 b,c, and c compare the atomistic (black) and CG (red) RDFs between CG type 1-2, 2-4, and 3-3. CG RDFs for those pairs replicate closely those from atomistic simulations.

The intermolecular potentials at three different temperatures are next derived. Multi state IBI (MS-IBI) is executed to find a single set of potentials that represent the target structure data at multiple states. SS-IBI finds the intermolecular potentials that the convergence percentage is 100%. However, in multi-state IBI, the matching percentage is only 60 % after 40 iterations. The convergence percentage increases as the iterative calculation is repeated, reaching 90 % when approximately 100 iterations are performed. After that, the percentage does not increase significantly, and fluctuates between the high 80s and the low 90s. It reaches a plateau after 200 iterations, and the maximum convergence is 95 percent. The second wells for interaction pair 1-3 and 1-4 do not converge. Thus, the set of intermolecular CG potentials are found, and the final values are listed in Table S.2.

We simulate the formation of lipid bilayers at different states using the selected CG parameters from MS-IBI. The selected temperatures are 298, 340, and 370 K, indicating the solid state, transition between the solid and liquid crystalline states, and liquid crystalline state of DSPE lipid bilayers, respectively. Figures 4.5 a, b, and c are snapshots of the lipid bilayer
formed at each temperature. Blue and dark yellow spheres indicate CG sites 1 and 2, which are ethanolamine and phosphate, respectively. Red and cyan colors represent ester linkages and hydrophobic tails. At 298K, DSPE lipid chains are well organized, and the tilted gel phase is observed, which is one of the solid state configurations of DSPE lipid bilayer. In addition, the very clear boundary between the upper and lower layers is identified. The snapshot of the lipid bilayer at the transition temperature shows a bilayer configuration that is very similar to that of solid state. The boundary between the upper and the lower layers is still observed, but the carbon tails inside the bilayer are irregular as opposed to arranged. Finally, at 370K, the arrangement of the head groups collapse, unlike the bilayers at 298 and 340 K. The hydrocarbon tails are also mixed together, resulting in overlap between the upper and lower layers. In addition, the interface between the upper and lower layers has vanished. Figures 4.5 d and e compare the atomistic and CG RDFs between sites 1-2 at three temperatures. As shown in the Figure 4.5d, the height of the first and second peaks in the atomistic model gradually decreases as the temperature increases. Similarly, in the case of CG RDF (Figure 4.5e), the height of the peak decreases as the temperature increases. Also, the decrease in the 2nd peak height due to the temperature change is well expressed. In conclusion, the simulated CG lipid bilayers found using the LIME intermolecular potential parameters from the MS-IBI method well represent the distinctive characteristics from atomistic simulations in the various states.

The bilayer thickness is measured at various temperatures. The bilayer thickness is defined to be the mean distance between phosphate atoms in the upper and lower layers. In the case of atomistic simulations, the thickness is directly determined according to the definition. However, in CG simulation, the thickness is measured using the locations of the CG sites containing phosphate. Figure 4.6 shows the DSPE lipid bilayer thickness in the atomistic
simulation (red) and in the CG simulations using parameters obtained from MS-IBI (black) at various temperatures. As shown in the figure, at low temperatures, the lipid bilayer thickness in the atomistic and CG simulations is high, reflecting the characteristic of the lipid bilayer in the solid state. As the temperature increases, the thickness gradually decreases due to the overlap of the upper and lower layers. Although the thickness variations of the bilayers obtained from the atomistic and CG simulations are very similar, a gap of about 2 Angstrom is found at all temperatures. This is because the CG phosphate is assumed to be located at the center of mass of the site, which is different from the position of the phosphate atom in the atomistic simulation. The thicknesses of CG lipid bilayers simulated with parameters obtained by SS-IBI (blue) are displayed, do not show comparable results with that of atomistic lipid bilayers over the temperature range, which reminds us of the necessity of using MS-IBI. Therefore, it is demonstrated that the set of potentials obtained through MS-IBI is suitable for the CG simulations of various temperatures.

4.4 Conclusion

We have developed an improved version LIME for DSPE lipids for use in coarse grained simulation. The DSPE lipid has 6 distinct coarse grained sites. Like the original LIME, the improved LIME extracts the parameters for the CG simulation from RDFs based on the atomistic simulation results. In the original LIME, the intermolecular interactions between CG sites are represented using a single square well potential, and a simple single-step Boltzmann inversion technique to extract the intermolecular potentials between CG sites. Those simplifications led to a discrepancy between the RDFs formed in CG simulations and those in
atomistic simulations. The improved model has multiple square wells which gives us a CG RDF that better reflects the atomistic RDF.

The original LIME extracted CG parameters from atomistic simulation results at a single temperature, which meant that the parameters were likely only valid at one temperature. CG-DMD simulations at a different temperature should in principle have different parameter sets. To overcome this drawback, we introduced a multistate IBI process to calculate a single set of intermolecular parameters that can be applied at various temperatures.

Using the parameters obtained through MS-IBI, the CG DSPE lipid bilayer was simulated at various temperatures from 298 to 370 K. The accuracy of the improved model is confirmed by comparison of the CG and atomistic RDFs, the bilayer morphology, and the bilayer thickness at various temperatures. Thus, the obtained parameters are well suited for use in systems where CG simulations should be performed at various temperatures, such as lipid bilayer phase behavior.

4.5 Acknowledgement

This work was supported by the National Institutes of Health (EB006006), the NSF’s Research Triangle MRSEC on Programmable Soft Matter, DMR-1121107 and CBET 1512059, and the NSF Partnership for Research and Education in Materials (PREM), DMR-1205670.
4.6 References


9. Lladó, V.; López, D. J.; Ibarguren, M.; Alonso, M.; Soriano, J. B.; Escribá, P. V.;


25. Matos, C.; Lima, J. L. C.; Reis, S.; Lopes, A.; Bastos, M., Interaction of


34. Marrink, S. J.; de Vries, A. H.; Mark, A. E., Coarse grained model for semiquantitative


Table S4.1. Minimum and maximum bond lengths for all bonded CG sites.

<table>
<thead>
<tr>
<th>CG site i</th>
<th>CG site j</th>
<th>Minimum (Å)</th>
<th>Maximum (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3.295</td>
<td>4.025</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.855</td>
<td>7.765</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>6.125</td>
<td>11.455</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>3.965</td>
<td>8.945</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3.895</td>
<td>4.885</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>7.165</td>
<td>8.545</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>9.615</td>
<td>12.215</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>4.535</td>
<td>6.265</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3.425</td>
<td>3.995</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>6.175</td>
<td>7.115</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8.905</td>
<td>11.465</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3.095</td>
<td>4.165</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>5.025</td>
<td>7.115</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3.245</td>
<td>3.885</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>6.245</td>
<td>7.655</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>9.165</td>
<td>11.415</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>4.505</td>
<td>7.215</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>3.305</td>
<td>3.885</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>6.285</td>
<td>7.655</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>9.165</td>
<td>11.415</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>3.295</td>
<td>3.885</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>6.245</td>
<td>7.655</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>8.475</td>
<td>10.815</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>3.255</td>
<td>3.885</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>5.645</td>
<td>7.055</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>2.765</td>
<td>3.275</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>3.225</td>
<td>3.555</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>6.145</td>
<td>7.705</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>8.415</td>
<td>10.615</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>3.575</td>
<td>4.285</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>5.895</td>
<td>7.235</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>9.185</td>
<td>11.415</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>2.645</td>
<td>3.035</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>5.905</td>
<td>7.235</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>8.795</td>
<td>10.995</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>3.575</td>
<td>4.295</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>6.505</td>
<td>8.075</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>8.735</td>
<td>11.215</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>3.255</td>
<td>3.875</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>5.635</td>
<td>7.045</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>2.765</td>
<td>3.275</td>
</tr>
</tbody>
</table>
Table S4.2. Interaction energy for each pair of CG type.

<table>
<thead>
<tr>
<th>CG type i</th>
<th>CG type j</th>
<th>$\varepsilon_1$ (eV)</th>
<th>$\varepsilon_2$ (eV)</th>
<th>$\varepsilon_3$ (eV)</th>
<th>$\varepsilon_4$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-0.57246</td>
<td>-0.27457</td>
<td>0.19260</td>
<td>-0.30884</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.37203</td>
<td>-0.75832</td>
<td>0.08382</td>
<td>-0.17206</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-0.96813</td>
<td>-0.34573</td>
<td>-0.06778</td>
<td>0.04122</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-0.36337</td>
<td>0.01094</td>
<td>0.35482</td>
<td>-0.06888</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-0.18169</td>
<td>-0.34831</td>
<td>-0.18062</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-0.53864</td>
<td>-0.11545</td>
<td>0.09997</td>
<td>-0.27255</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-0.52639</td>
<td>-0.29586</td>
<td>0.28110</td>
<td>-0.05991</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>-0.40208</td>
<td>-0.72231</td>
<td>-0.35277</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-0.14873</td>
<td>-0.44539</td>
<td>-0.40409</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>-0.56751</td>
<td>-1.23612</td>
<td>-0.55366</td>
<td>-0.42000</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>-0.91002</td>
<td>-0.12000</td>
<td>0.37000</td>
<td>-0.08000</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>-0.06340</td>
<td>-0.64551</td>
<td>-0.37375</td>
<td>-0.64120</td>
</tr>
</tbody>
</table>
Figure 4.1. (a) All-atom, (b) united-atom, and (c) coarse-grained representation of DSPE. The color scheme is: blue (type1 - ethanolamine); yellow (type2 - phosphate group); red and orange (type3 and 4 - ester linkage); cyan (type5 - alkyl tail group); green (type6 - terminal tail group).
Figure 4.2. Plot comparing the continuous potential energy and our discontinuous potential energy between coarse-grained site 1 and 2.
Figure 4.3. (a) The convergence percentage of $g_{CG}(r)$ and $g_A(r)$ in SS-IBI. Comparison atomistic (black) and CG (red) RDFs between CG sites of types (b) 1-2, (b) 2-4, and (c) 3-3.
Figure 4.4. The convergence percentage of $g_{CG}(r)$ and $g_{A}(r)$ in SS-IBI and MS-IBI.
Figure 4.5. Snapshots of a DSPE lipid bilayer in CG DMD simulation at (a) $T = 298\, \text{K}$, (b) $T = 340\, \text{K}$, and (c) $T = 370\, \text{K}$. RDFs from (d) atomistic and (e) CG simulations at three temperatures.
Figure 4.6. Comparison of the bilayer thickness measured from (a) the atomistic simulations (red), CG simulations using the (b) SS-IBI (blue) and (c) MS-IBI (black) intermolecular potentials.
Chapter 5

Future Work

In this work we have described the development of coarse-grained model for hydrogel and assembly and for lipid bilayer assembly that could be used for drug delivery applications. We discuss possible future directions based upon our research.

5.1 Investigation of DNA-mediated hydrogel degradation when it encounters target molecules

Nucleic acids, and more specifically aptamers, can be incorporated as structural and functional components of hydrogels. We can take advantage of their ability to identify and interact with molecular targets and subsequently trigger hydrogel degradation and consequently release drugs. Our collaborator, Dr. Betancourt synthesizes discrete aptamer-crosslinked PEG-based hydrogel particles that respond to the molecular target adenosine. Although we hypothesize that interaction of the hydrogels with the target molecules will result in rapid disintegration of crosslinks due to preferential binding of the aptamer to its target, the detailed binding mechanism of the adenosine is still unknown. In a future project, we will develop a model of a target molecule (adenosine) that triggers the hydrogel to degrade and consequently release drugs upon exposure to a specific oligonucleotide (aptamer) sequence. These studies could be useful to evaluate the potential of these hydrogels as molecularly-controlled drug delivery systems.
5.2 Investigation of polyethylene grafts on lipid bilayer surfaces for the purpose of protecting liposomes.

Surface modification of liposomes with grafted Polyethylene glycol (PEG) chains creates an aqueous layer surrounding the liposome surface. The aqueous layer width is called the fixed aqueous layer thickness. The water molecules around the PEG chains act as a protective barrier against the uptake of liposomes by reticuloendothelial cells. Liposomes having PEGylated chains on their surface can avoid being caught by the immune system of humans, significantly increasing their survival time. Thus, it is essential to have a thick fixed aqueous layer thickness to maximize the PEG’s protecting function and circulation time. In the future we would like to use our multi-scale modeling approach to extract parameters for PEG from atomistic simulations. This would allow us to run simulations to try to understand the conformation of grafted PEGs on the lipid bilayer and to find the conditions for maximizing the size of the fixed aqueous layer. The conditions that affect the thickness are molecular weight of PEG, grafting density on the liposome surface, and salt concentration in solution.