ABSTRACT

LI, NAN. Molecular Modeling of the Self-assembly of Stimuli-responsive Biomacromolecules. (Under the direction of Dr. Yaroslava G. Yingling).

Self-assembly processes of stimuli-responsive amphiphilic macromolecules are ubiquitous in industrial and biological processes; understanding their physical properties can also provide insights into the design of soft materials with novel and tailored properties. For example, elastin-like polypeptides (ELPs) exhibit lower critical solution temperature (LCST) phase behaviors in aqueous solution, which undergo temperature-triggered coacervation upon heating above the cloud point temperature. As another example, in response to a change of salt concentration, the self-assemblies of polyelectrolyte block copolymers undergo morphological transition from cylindrical micelles to vesicles. The mechanism that controls the self-assembly and stimuli-responsive behaviors is important, yet not fully understood.

Computational techniques offer a range of advantages in elucidating the responsive mechanism in various processes. Rapid advances in computational resources and the development of novel algorithms enable molecular simulations to address larger systems and tackle the critical issues in the area of soft materials. Here, various modeling techniques are used to predict and explain the stimuli-responsive behavior of biomacromolecules. This dissertation examines two cases of responsive materials: (1) the molecular origin of LCST behavior of ELPs and (2) the predictions of solvent-induced morphological changes of polyelectrolyte block copolymers, utilizing molecular modeling techniques.

In the first case, the effect of temperature on the structure, dynamics and association of ELPs poly(VPGVG)n and poly(VGPVG)n in aqueous solution is examined using atomistic molecular dynamics simulations. The study highlighted the role of temperature and the amino acid order in conformational properties, hydrophobicity, local residue interactions,
and the secondary structure propensity of ELPs. We conclude that the LCST phase behavior of poly(VPGVG) is a collective phenomenon that originates from the correlated gradual changes in single polypeptide structure and the abrupt change in properties of hydration water around the peptide and is a result of a competition between peptide–peptide and peptide–water interactions. The comparison study between poly(VPGVG) and poly(VGPVG) at a range of temperatures leads us to propose that conformational preferences of the constituent amino acids in a specific order play an important role in their structure, dynamics and thermodynamics properties.

In the second case, the self-assembly properties of polyelectrolyte block copolymers in response to solvent ionic strength is examined. Modeling of complex polyelectrolyte systems can be challenging and computationally intensive due to the implementation of long range electrostatic interactions. Therefore, we developed a new coarse-grained methodology for the modeling and simulation of polyelectrolyte systems with implicit representation of ionic strength in the solvent (ISIS) for dissipative particle dynamics (DPD) simulations. In this study, we systematically analyzed how the ionic strength of the solution and the length of the polyelectrolyte block affect the kinetics of self-assembly and the equilibrium morphology of the aggregates. With the utilization of our ISIS DPD model, a comprehensive set of data was obtained to construct the phase diagram of amphiphilic polyelectrolyte diblock/triblock copolymers in aqueous solution. Various morphologies were obtained, such as micelles, vesicles, lamellar aggregates and micellar networks. Quantitative evaluation of the micelle radius of gyration, radius of the micellar core, corona thickness for diblocks and polyelectrolyte bridge fraction for triblocks were conducted. Their scaling law dependence on the solvent ionic strength, or the length of the polyelectrolyte block was obtained.
The results of this study may aid in the design of effective polymeric or protein-based materials in a number of promising applications such as drug and gene delivery systems, tissue engineering, biosensors, oil and water clean-up and recovery and numerous nanoscale devices.
Molecular Modeling of the Self-assembly of Stimuli-responsive Biomacromolecules

by
Nan Li

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

Materials Science and Engineering

Raleigh, North Carolina

2016

APPROVED BY:

_______________________________
Dr. Yaroslava G. Yingling
Committee Chair

_______________________________
Dr. Donald Brenner

_______________________________
Dr. Albena Ivanisevic

_______________________________
Dr. Carol K. Hall

_______________________________
Dr. Stefan Zauscher
DEDICATION

To my parents.

致我的父母。
BIOGRAPHY

Nan Li was born to Yulan Zhou and Jianbo Li in Xingtai, Hebei, China. She graduated from Jilin University with a B.S in Mechanical Engineering in 2008. She obtained her M.S in Mechanical Design and Theory in 2011 and conducted her thesis research under the supervision of Dr. Chuncheng Zuo. She subsequently started as a PhD student in the Department of Materials Science and Engineering at NCSU, under the supervision of Dr. Yaroslava G. Yingling.
ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my supervisor, Dr. Yaroslava G. Yingling, who has guided and supported me throughout my PhD work. I am very grateful that there has always been a hand to grab as I walk through this academic path.

I would like to thank my dissertation committee. Dr. Donald Brenner, Dr. Carol K. Hall, Dr. Albena Ivanisevic, and Dr. Stefan Zauscher have each provided helpful feedback and have been great mentors who have prepared me to get to this milestone in my academic life.

I must acknowledge all of my labmates for their help, advice and insight throughout my stay in the lab. I also would like to express my gratitude to all of my friends whose friendships have made my stay in Raleigh enjoyable.

I would also like to thank my family. My parents and my husband, Dr. Kai Feng, have been extremely supportive throughout all my studies. Thank you.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... ix
LIST OF FIGURES ......................................................................................................... x

Chapter 1 Introduction ................................................................................................... 1
  1.1 Background ........................................................................................................... 1
  1.2 Biomaterials ....................................................................................................... 4
    1.2.1 Elastin-like polypeptide (ELP) ...................................................................... 4
    1.2.2 Deoxyribonucleic acid (DNA) ..................................................................... 6
    1.2.3 DNA block copolymers .............................................................................. 8
  1.3 Methods ............................................................................................................... 9
    1.3.1 Molecular Dynamics .................................................................................. 9
    1.3.2 Dissipative Particle Dynamics ................................................................... 12
  1.4 LCST behavior of polymers and biopolymers ................................................... 14
  1.5 Self-assembly of amphiphiles ............................................................................ 16
  1.6 References ......................................................................................................... 18

Chapter 2 Progress in molecular modeling of DNA materials ........................................ 30
  2.1 Introduction ....................................................................................................... 30
  2.2 Molecular Modeling .......................................................................................... 33
    2.2.1 Atomistic simulation ................................................................................. 33
    2.2.2 Coarse-grained model ............................................................................. 35
  2.3 Modeling DNA-based materials ......................................................................... 37
    2.3.1 DNA-nanoparticle assemblies ................................................................. 37
    2.3.2 DNA functionalized on gold surface ........................................................ 39
    2.3.3 Graphene and DNA ................................................................................ 40
    2.3.4 Nonspecific DNA and nanoparticle assemblies ....................................... 42
    2.3.5 DNA origami ............................................................................................ 43
  2.4 Conclusions ....................................................................................................... 44
  2.5 References ......................................................................................................... 45

Chapter 3 Molecular description of the LCST behavior of an elastin-like polypeptide .... 54
  3.1 Introduction ....................................................................................................... 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Materials and Methods</td>
<td>60</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Single ELP Simulation</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Interactions between two ELPs</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>64</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Conformation of a single (VPGVG)\textsubscript{18} polypeptide</td>
<td>64</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Secondary structure analysis</td>
<td>66</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Structure of hydration water</td>
<td>70</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Solvent accessible surface area and energy</td>
<td>73</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Polypeptide aggregation</td>
<td>75</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusions</td>
<td>77</td>
</tr>
<tr>
<td>3.5</td>
<td>References</td>
<td>81</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>LCST Behavior is Manifested in a Single Molecule: Elastin-like polypeptide (VPGVG)\textsubscript{n}</td>
<td>95</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>Materials and Methods</td>
<td>101</td>
</tr>
<tr>
<td>4.2.1</td>
<td>ELP simulations with single chain.</td>
<td>101</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Hydrogen bonding analysis.</td>
<td>102</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Secondary structure analysis and interaction energy calculation.</td>
<td>102</td>
</tr>
<tr>
<td>4.3</td>
<td>Results</td>
<td>103</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion</td>
<td>111</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusion</td>
<td>114</td>
</tr>
<tr>
<td>4.6</td>
<td>References</td>
<td>117</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>The Role of Amino Acid Order in Temperature-responsive Properties of Elastin-like Polypeptides: Molecular Dynamics Simulations of (VPGVG)\textsubscript{18} and (VGPVG)\textsubscript{18}</td>
<td>128</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>128</td>
</tr>
<tr>
<td>5.2</td>
<td>Materials and Methods</td>
<td>130</td>
</tr>
<tr>
<td>5.3</td>
<td>Results</td>
<td>132</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Structural properties and water-peptide interactions</td>
<td>132</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Dihedral angles, secondary structures and role of individual residue</td>
<td>134</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Hydration water network</td>
<td>140</td>
</tr>
</tbody>
</table>
9.3 Materials and Methods ........................................................................................................ 249
9.4 Conclusion .......................................................................................................................... 252
9.5 References ........................................................................................................................ 253

Chapter 10 Enzymatic Polymerization of High Molecular Weight DNA Amphiphiles that Self-Assemble into Star-like Micelles ............................................................................. 264
10.1 Introduction ....................................................................................................................... 264
10.2 Experimental Section ....................................................................................................... 265
10.3 Results ............................................................................................................................... 268
10.4 References ....................................................................................................................... 278
APPENDIX ............................................................................................................................... 287
LIST OF TABLES

Table 4.1  Summarizing properties of ELPs at lengths (L=10, 18 and 30) averaged over temperatures below $T_t$, above $T_t$ and the difference between them……. 123

Table 5.1  Distance between $\text{Ca}$@Val1-$\text{Ca}$@Val4 in pentamer (VPGVG) and (VGPVG) at $290 \text{ K}$ and $350 \text{ K}$, respectively……………………………………. 151

Table 5.S1  Dihedral angles obtained from the highest intensity points in Ramachandran plots in Figure 5.S6……………………………………………………… 152

Table 5.S2(a)  Secondary structure propensities for (VPGVG)$_{18}$ from simulation at various temperatures………………………………………………………… 153

Table 5.S2(b)  Secondary structure propensities for (VGPVG)$_{18}$ from simulation at various temperatures………………………………………………………… 154

Table 5.S3  (a)The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VPGVG at 290 K (b)The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VPGVG at 350 K (c)The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VGPVG at 290 K (d)The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VGPVG at 350 K……………………………………………………………………………… 155

Table 6.1  Secondary structure propensities for (VPGVG)$_{18}$ at 0 ns and 700 ns…… 174

Table 6.2  Secondary structure propensities for (VGPVG)$_{18}$ at 0 ns and 700 ns…… 174

Table 8.1  Dependence of the scaling exponents for $H_{\text{corona}}$ and $R_{g,m}$ on the repulsive parameter $a_{pp}$ and solvent ionic strength $c_s$. (represented by the electrostatic repulsive parameter $a_{elec}$, $c_s \sim a_{elec}^{-1}$)……………………………………………… 227

Table 8.2  The dependence of the scaling exponents for $R_{g,m}$ and $H_{\text{corona}}$ on the polyelectrolyte length $N_A$………………………………………………………… 227

Table 8.S1  Micellar size from AFM images at different salt concentrations……………… 227

Table 8.S2  Experimental results…………………………………………………………………………………………………… 228

Table 10.1  Average hydrodynamic radii ($R_h$) and micellar core heights ($H_{\text{core}}$) obtained from DLS measurements and AFM images, respectively……… 282
LIST OF FIGURES

Figure 3.1  (a) Side and front view of the β-spiral structure for (VPGVG)18. Final simulation snapshots of (b,c) a single (VPGVG)18 structure and (d,e) the interaction between two ELPs at (b,d) 290 K and (c,e) 350 K. The backbone is represented as secondary structure ribbons where turn motif is colored in green, β-sheet is yellow and coil is silver. All heavy atoms are depicted as sticks. In panel a-c, the side chains are colored by amino acid (Gly in purple, Pro in yellow and Val in blue). In panel d-e, the side chains are colored by molecule……………………………………………………………………….. 87

Figure 3.2  (a) Temperature dependence of the radius of gyration (Rg) of (VPGVG)18. (b) Correlation between Nw, the number of water molecules in the second watershell of backbone, and npw, the number of peptide-water hydrogen bonds. The temperature is labeled at the data points. (c) Number of the intramolecular intra-peptide hydrogen bonds (black) and intramolecular water-mediated peptide hydrogen bonds (red). The error bars in these plots represent the standard deviation……………………………………………. 88

Figure 3.3  Secondary structure formation as a function of temperature for (a) β-turn and (b) β-strand structures…………………………………………………………………………………………………………………………… 89

Figure 3.4  Ramachandran plot (φ, ψ distributions) for each residue in a pentamer, (a1,2) Val1, (b1,2) Pro2, (c1,2) Gly3, (d1,2) Val4, (e1,2) Gly5, at two temperatures, 290 K (left column) and 350 K (right column), colored by intensities. The most populated regions on the Ramachandran plots are labelled…………… 90

Figure 3.5  (a) Radial distribution function (RDF) of oxygen atoms in water molecules around polypeptide atoms as a function of temperature. (b) Probability distribution of the size Nmax of the largest water network around polypeptide. (c) Probability distribution of the size N_{max}^W of the largest water network in pure water systems…………………………………………………………………………………. 91

Figure 3.6  Temperature dependence of properties of a single ELP chain: (a) Interaction energy of each individual residue in an ELP pentamer with water. (b) Fraction of SASA of Val1_side chain (black), Pro2_side chain (blue), Val4_side chain (red), Gly3 (green) and Gly5 (pink). (c) SASA of peptides (black; square) and SASA fraction of Val1_side chain (blue; triangle) and Pro2_side chain (red; circle)………………………………………………………………………………………………………………… 92

Figure 3.7  Interaction between two ELPs. (a) Distance between center-of-mass of two polypeptides and (b) peptide-peptide interaction energy at 290 K and 350 K…………………………………………………………………………………………………………………………………………………………………………………………………………………. 93
Figure 3.8 Peptide-water interaction energy in single-peptide system (black; square). Peptide-peptide interaction energy in double-peptide system (blue; triangle). The data are averaged over the last 40 ns for each MD run in a single chain system and over the last 5 ns run for double chain system. 94

Figure 4.1 Number of hydrogen bonds between peptide and water, Npw, versus number of water molecules within the first hydration shell of peptide backbone, Nw. (A) (VPGVG)30, the points are grouped into two separate clusters suggesting that Tt is between 305 K and 310 K. (B) (VPGVG)18 with Tt between 330 K and 335 K. (C) (VPGVG)10 with Tt between 340 K and 350 K also has two distinct clusters. (D) GVG(VPGVG)3 does not have distinct clusters. Clusters below Tt are colored red, clusters above Tt are colored blue. The standard deviation is based on the time average of each simulation. 124

Figure 4.2 Transition temperature versus chain length from experiment (blue squares) and simulation (red triangles) based on Npw-Nw analysis. Both data sets are fitted with power function Tt = c*(Length)^ν. The coefficient from experiments is Ce = 473.81, and from simulations is Cs = 354.93. The exponent from experiments is -0.63, and from simulations is -0.65. The simulation Tt value for (GVG)(VPGVG)3 (blue empty square), 157 °C, is obtained by extrapolating the fitted power function to the length 3.5. 125

Figure 4.3 Number of H-bonds between (A) peptide and water, Npw, and (B) peptide and peptide, Npp, versus temperature. Slopes of linear fit of all the curves are shown next to each curve. Inset graphs depict the average values of the slope per pentamer for different lengths. 126

Figure 4.4 (A) Difference between the average of npw or npp over all the temperatures above Tt and the average of npw or npp over all the temperatures below Tt for each pentamer, Δnpw = npw_above - npw_below, Δnpp = npp_above - npp_below. (B) Schematic representation of the configurations of ELP at temperatures (a) below Tt and (b) above Tt. The peptide backbone is grey, peptide-water H-bonds are green and peptide-peptide H-bonds are magenta. Water molecules are represented by 3 circles; circles in red are oxygen atoms and in blue are hydrogen atoms. 127

Figure 5.1 Temperature dependence of properties of a single ELP chain: (a) radius of gyration (Rg) of (VPGVG)18 and (VGPVG)18 (b) npw, the number of peptide-water hydrogen bonds. (c) solvent accessible surface area (SASA). (d) Peptide-water interaction energy. The error bars in these plots represent the standard deviation. The same representations are used in the following figures. 156
Figure 5.2 (a-d) Ramachandran plot ($\phi$, $\psi$ distributions) for Val1 and Val4 in (VPGVG)18 and (VGPVG)18, respectively, at 350 K, colored by intensities. The most populated regions on the Ramachandran plots are labelled. (e-f) Representative structures for G(VPGVG)V segment at 350 K (e) and G(VGPVG)V segment at 350 K (b) built with the dihedral angles obtained from the highest intensity points in Ramachandran plots (Figure 5.3 and Figure 5.S3) ................................................................. 157

Figure 5.3 Temperature dependence of properties of a single ELP chain: (a) Interaction energy of each residue in an ELP pentamer with water. (b) Number of hydrogen bonds formed between each residue in an ELP pentamer and water. (c) SASA of each residue in an ELP pentamer ......................................................... 158

Figure 5.4 (a) Probability distribution of the size Nmax of the largest water network around polypeptide. (b) A snapshot of (VGPVG)18 and the largest water network around it at 350 K ................................................................. 159

Figure 5.5 Interaction between two ELPs at 350 K. (a) Distance between center-of-mass of two polypeptides and (b) peptide-peptide interaction energy (c) Number of inter-chain hydrogen bonds ......................................................... 160

Figure 5.S1 Temporal evolution of radius of gyration (Rg) of (VPGVG)18 at different temperatures. Representative structures of (VPGVG)18 from cluster analysis of MD trajectories at (a-c) 290 K and (d-f) 350 K are displayed .......... 161

Figure 5.S2 Temporal evolution of radius of gyration (Rg) of (VGPVG)18 at different temperatures. Representative structures of (VGPVG)18 from cluster analysis of MD trajectories at (a-c) 290 K and (d-f) 350 K are displayed .......... 162

Figure 5.S3 Time autocorrelation function of (VPGVG)18 backbone Ca- Ca bond vectors: (a) unit vector along backbone Ca- Ca bond; (b) unit vector out of plane. We can estimate the relaxation time of (VPGVG)18 backbone vectors from the time autocorrelation profiles which is less than 25 ns for each group ................................................................. 163

Figure 5.S4 Time autocorrelation function of (VPGVG)18 peptide backbone Ca- Ca bond vectors: (a) unit vector along backbone Ca- Ca bond; (b) unit vector out of plane. We can estimate the relaxation time of (VPGVG)18 backbone vectors from the time autocorrelation profiles which is less than 15 ns for each group ................................................................. 164

Figure 5.S5 (a) Temperature dependence of Number of the intramolecular intra-peptide hydrogen bonds (np) and intramolecular water-mediated peptide hydrogen bonds (npwp) of (VPGVG)18 and (VGPVG)18. (b) Temperature dependence of the number of water molecules in the first watershell of (VPGVG)18 and (VGPVG)18 ................................................................. 165
Figure 5.S6  Ramachandran plot ($\phi$, $\psi$ distributions) for each residue in a pentamer, (a) VPGVG at 290 K, (b) VPGVG at 350 K, (c) VGPVG at 290 K, (d) VGPVG at 350 K, colored by intensities. The most populated regions on the Ramachandran plots are labeled…………………………………………. 166

Figure 5.S7  Average frequency of occurrence of each residues in (VPGVG)$_{18}$ (a, b) and (VGPVG)$_{18}$ (c, d) in each type of secondary structural motifs at 290 K (a, c) and 350 K (b, d)……………………………………………………………. 167

Figure 6.1  Snapshots of 27 (VPGVG)$_{18}$ at (a) 0 ns and (b) 700 ns, and the snapshot of 27 (VGPVG)$_{18}$ at (c) 700 ns. (d) Number of polymer clusters in the simulation system as a function of simulation time…………………………... 175

Figure 6.2  Temporal evolution of (a) the averaged radius of gyration of polypeptides and (b) the averaged number of intra-chain hydrogen bonds of polypeptides………………………………………………………………………… 176

Figure 6.3  Number of inter-chain hydrogen bonds as a function of simulation time... 177

Figure 6.4  Ramachandran plot ($\phi$, $\psi$ distributions) for each residue in repeated pentamer in (a) (VPGVG)$_{18}$ at 0 ns, (b) (VPGVG)$_{18}$ at 700 ns, (c) (VGPVG)$_{18}$ at 0 ns, and (d) (VGPVG)$_{18}$ at 700 ns, colored by intensities……………..…… 178

Figure 7.1  Properties of a single long polyelectrolyte chain as a function of solvent ionic strength. (a) End-to-end distance as a function of the repulsive parameter, $a_{pp}$; (b) Logarithmic plot of the end-to-end distance as a function of $a_{elec}$; (c) Radius of gyration as a function of $a_{elec}$; and (d) Dependence of the chain persistence length, $l_p$, on the electrostatic repulsive parameter, $a_{elec}$. Note that $a_{elec} \sim c_s^{-1}$……………………………………………………… 195

Figure 7.2  Snapshots of micellization of polyelektrolyte copolymers in DPD simulations. (a-d) Timeline of a 300-chain system with $a_{pp} = 40$. (a) Time step =1, (b) Time step = 1.0E5(c) Time step = 1.0E6, and (d) Time step =4.0E7. (e-h) Typical snapshots of polyelectrolyte block copolymers in aqueous solution at (e) $a_{pp} = 25$, (f) $a_{pp} =30$, (g) $a_{pp} =50$, and (h) $a_{pp} =60$. The yellow and blue beads represent the hydrophobic block and hydrophilic block, respectively. Water is omitted for clarity.................................................. 196

Figure 7.3  The properties of polyelectrolyte diblock micelles as a function of solvent ionic strenght. (a) Probability distribution of the anisotropy of the micellar core and (b) Probability distribution of the aggregation number................. 197

Figure 8.1  (a) Representation of a PDC chain. Simulation snapshots of initial (b) and final configurations at (c) $a_{pp}= 25$ and (d) $a_{pp}= 40$ with $N_A= 50$ and $N_B= 4$. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively. (e) Phase diagrams of typical aggregates of PDCs obtained by
DPD simulations. The section views for vesicles and side views for lamellar shapes are marked by the dashed circles and ellipsoids. The phase diagrams of final snapshots of PDCs in aqueous solutions are displayed in Figure 8.S1………………………………………………………………………. 229

Figure 8.2 Properties of micellar aggregates and morphological transitions represented as contour plots for anisotropy indices of (a) the micellar hydrophobic core, \( \kappa_c^2 \) and (b) the whole micelle, \( \kappa_m^2 \), and (c) the aggregation number, \( P \). The region with aggregates formed by all chains in the system is shadowed……………………………………………………………………….………………….. 230

Figure 8.3 The thickness of corona (\( H_{\text{corona}} \)) of spherical PDC micelles as a function of (a1) repulsive parameter \( a_{pp} \), (a2) length of polyelectrolyte block \( N_A \), and (a3) \( a_{elec} \). The radius of gyration (\( R_{g,m} \)) of micelles as a function of \( a_{pp} \) (b1), \( N_A \) (b2), and \( a_{elec} \) (b3). Electrostatic repulsive parameter \( a_{elec}^{-1} c_s^{-1} \)………………………………………………………………….……………………….. 231

Figure 8.4 (a) Radius of gyration for F-dUTP-b-polyT50 micelles determined from SLS (black squares) and from DPD simulations (blue triangles) plotted as a function of the charge ratio \( \lambda \) (bottom) and repulsive parameter \( a_{pp} \) (top). The vertical dashed line indicates the predicted change in micellar morphology. (b-d) AFM Tapping Mode height images of spherical micelles and aggregates at three different charge ratios………………………… 232

Figure 8.S1 a Phase diagrams of dilute aqueous solution of polyelectrolyte diblock copolymers obtained by DPD simulations. Both hydrophobic and polyelectrolyte beads are displayed in (a) and only hydrophobic beads are displayed in (b). Aggregates formed by all 300 chains, including vesicles, lamellar shape or pretzel shape are outlined with a black box, wormlike/cylindrical micelles with a red box, and spherical micelles with a blue box. The section view for vesicles and side view for lamellar shapes are displayed in dash circle………………………………………………………………………. 233

Figure 8.S2 Typical snapshots of polyelectrolyte block copolymers in aqueous solution at \( N_A=4 \). Total number of chains is 300 (a1, a2, a3 and a4) and 600 (b1, b2, b3 and b4)………………………………………………………………………………………… 236

Figure 8.S3 Temporal evolution of averaged aggregation number \( P \) for (a) \( N_A=30 \) and (b) \( N_A=50 \)……………………………………………………………………………….. 237

Figure 8.S4 Probability distribution of the aggregation number when \( N_A=50 \)………… 238

Figure 8.S5 Radius of gyration of the micelles \( R_{g,m} \) from DPD simulations as a function of repulsive parameter \( a_{pp} \)…………………………………………………………………………………………. 239
Figure 8.S6 Contour plot of ratio of the thickness of corona $H_{corona}$ to the radius of core $R_{g,c}$. The region with aggregates formed by all chains is shadowed. 240

Figure 8.S7 The radius of gyration of micellar core as a function of $a_{pp}$ (a) and $N_A$ (b). 241

Figure 9.1 (a) Representation of a polyelectrolyte triblock chain. (b) Schematic illustration of typical loop and bridge configurations for middle block. (c-e) The phase diagrams of final snapshots of PDCs in aqueous solutions obtained by DPD simulations at the polymer volume fraction of 10%, 15% and 20%, respectively. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively. 257

Figure 9.2 The percentage of bridge chains (a) and the number of hydrophobic cores (b) as a function of repulsive parameter $a_{pp}$ for $N_A=30$ and $N_B=4$ at polymer volume fraction of 10%. 258

Figure 9.3 The averaged aggregation number as a function of repulsive parameter $a_{pp}$ at polymer volume fraction of 10%. 259

Figure 9.4 (a-c) The percentage of bridge chains as a function of $a_{elec}$ at the polymer volume fraction of 10%, 15% and 20%, respectively. (d) The exponent factor $\xi$ was plotted as a function of polyelectrolyte length. 260

Figure 9.S1 Different morphologies obtained with variation of the repulsive parameter $a_{pp}$ for $N_A=50$ and $N_B=4$ at polymer volume fraction of 10%. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively. 261

Figure 9.S2 The averaged aggregation number as a function of repulsive parameter $a_{pp}$ for (a) $N_A=50$, (b) $N_A=70$, and (c) $N_A=90$. 262

Figure 9.S3 The number of hydrophobic cores as a function of repulsive parameter $a_{pp}$. 263

Figure 10.1 Schematic, showing a) the enzyme-catalyzed synthesis of diblock polynucleotide amphiphiles, and b) their self-assembly into star like micelles. 283

Figure 10.2 Agarose gel electrophoresis results showing a) enzymatically synthesized homo-polynucleotides (poly(dTTP)) with three different MWs, using M:I of 200:1, 500:1, and 1000:1 (lanes pT1-pT3), and b) the same three poly(dTTP) enzymatically end-functionalized with B-dUTP (lanes pT1B-pT3B), at a molar ratio of B-dUTP to poly(dTTP) of 10:1. The presence of B-dUTP was verified by BODIPY fluorescence. 284
Figure 10.3 AFM TappingMode height images of star like micelles (pT1B, pT2B and pT3B), and the corresponding histograms of the measured micellar core heights. Inset: 3-D images of typical micelles (scale bar: 50 nm XY, 1.5 nm Z) g-h) Snapshots of equilibrium ssDNA micelles from the DPD simulations trajectory, where yellow beads represent the hydrophobes and blue beads represent hydrophilic polynucleotides .................................................. 285

Figure 10.4 Micelle $R_g$ at fixed aggregation number ($m = 5$) from DPD simulations, and the $R_h$ for micelles from DLS, plotted against hydrophilic block length ($N_A$). The slope of the lines through the data points reflects the power in the scaling law relationships .......................................................... 286
Chapter 1 Introduction

1.1 Background

In nature numerous supramolecular structures of varying complexity are formed via self-assembly of biomacromolecules in aqueous media. Many of such biomacromolecules are amphiphilic, i.e. they consist of hydrophilic (e.g. amide groups, ether groups) and hydrophobic moieties (e.g. vinyl backbone). The hydrophilic groups are able to interact strongly with water molecules and ensure water solubility. The hydrophobic group gives rise to a strong solvent mediated “attraction” between these hydrophobic groups to minimize the contact surface between hydrophobes and water.\(^1\) The delicate balance between hydrophilic and hydrophobic effects determines the solubility of polymers in water and the morphology formation. For polymer solutions, a decrease of the solvent thermodynamic quality tends to decrease polymer–solvent interactions and increase polymer–polymer interactions. One can induce self-assembly or disassembly by adjusting the solvent quality in the polymeric systems with stimuli-responsive solute. The specific properties of stimuli-responsive polymers can be tuned by certain stimulus, such as temperature, pH, electric and magnetic fields or mechanical stress. Some water-soluble polymers undergo phase separation upon heating, which can be described by a phase diagram with an LCST. The LCST behavior is usually detected by turbidity measurement, which precisely describes the macroscopic transition. However, their nature might be very subtle and their detection is not as easy as the detection of macroscopic phase transition.
The coacervation of tropoelastin plays a crucial role in the assembly into elastic fibers. This coacervation is based on the LCST behavior of tropoelastin, which causes tropoelastin structural self-assembly upon raising the temperature. The most interesting characteristic of tropoelastin is its ability to self-assemble under physiological conditions, thereby demonstrating an inverse temperature transition or lower critical solution temperature (LCST) transition phase behavior in aqueous solutions, i.e., they undergo a first-order phase transition into a polypeptide-rich and a water-rich phases upon heating above temperature cloud point temperature (T_{LCST}) \(^2\). Elastin-like polypeptides (ELP), are peptides polymers inspired by elastin polymers, which are defined by a pseudo-periodic, low complexity sequence with several types of repeat motifs. The lower critical solution temperature T_{LCST} of ELPs is dependent on pH\(^3\), salt\(^4\), polypeptide hydrophobicity\(^5\), and length\(^6\). The stimulus-responsive character of ELPs has led to their use in a variety of applications, including stimuli-triggered molecular actuators for recombinant protein purification\(^8\), drug delivery\(^9\), and as stimuli-responsive materials for tissue engineering.\(^10\) However, despite the proliferation of applications of ELPs in recent years, the origin of the LCST behavior of ELPs is still a matter of controversy\(^11\), and a detailed understanding is needed to enable further progress into the development of new elastin-based materials.

Synthetic amphiphilic block copolymers also form self-assembled structures in aqueous media. Their ultimate self-organization is determined by the characteristically complex equilibrium of noncovalent forces such as electrostatic, steric, hydrogen bonding, Van der Waals, and hydrophobic. Polyelectrolyte block copolymers, which combine the
properties of polyelectrolytes (i.e., sensitive to changes in solvent ionic strength and pH) with those of surfactants, may self-assemble into a variety of nanoaggregates in aqueous environment, such as micelles, vesicles and lamellar mesophases or micellar aggregates. For example, DNA block copolymers consist of synthetic oligonucleotides (ODN)s and hydrophobic segment maintain the special features of the biomacromolecule DNA and polymeric block type architectures that have attractive material properties. Micelles composed of these materials exhibit a corona of single-stranded (ss) DNA and have been utilized for the delivery of antisense ODNs\textsuperscript{12} and cancer drugs\textsuperscript{13}, for the hybridization with DNA-coated gold nanoparticles and as programmable\textsuperscript{14}, three-dimensional (3D) scaffolds for DNA-templated organic reactions\textsuperscript{15}. Furthermore, the self-assembly behavior of polyelectrolyte block copolymers can be efficiently tuned by variations in the polymer architecture, block length, ionic strength or/and pH in the aqueous solution. These qualities make the material properties of polyelectrolyte block copolymers and their self-assembly easy to control through the appropriate selection of polyelectrolyte block and neutral polymer block as well as the environmental factors.

Experimental methods for detailed studying of dynamics and structure of responsive biomacromolecules, such as nucleic acids, peptide, and biomolecules-functionalized materials are yet to be fully developed. Therefore, it is useful to develop and employ a complimentary approach to boost the development of potential applications of stimuli-responsive biomaterials. Molecular simulation techniques, which allow for direct visualization of the processes and interactions involved at molecular scale, have proven to be
essential in elucidating dynamical processes in various systems. Rapid advances in computational resources and algorithms enable molecular simulations to address complex and large systems and tackle the fundamental issues in the area of biomaterials.

1.2 Biomaterials

1.2.1 Elastin-like polypeptide (ELP)

This section introduces basic concepts of the molecular architecture and overall properties of ELP that are important for understanding the compositions and design of elastin-inspired LCST biopolymers. Elastomeric proteins are a class of biopolymers which are characterized by rubber-like elasticity, large extensibility before rupture, reversible deformation without loss of energy, and high resilience upon stretching. Elastin is a one natural type of elastomeric protein which is responsible for the extensibility and elastic recoil exhibited by many vertebrate tissues, including skin, lungs, and larger blood vessels. Despite the great diversity in elastin’s structures, some common features seem to have been preserved in elastin during evolution, such as the presence of significant amounts of glycine, proline and several aliphatic residues. The repetitive nature of elastin was firstly noticed through the studies conducted by Sandberg on elastin from swine. In their study, the repeats of the tetra-mer VPGG, penta-mer VPGVG and the hexa-mer VAPGVG in tryptic digests of procine elastin were identified.

Tropoelastin is the soluble precursor of elastin and consists of alternating hydrophobic and hydrophilic peptide domains. The most common amino acids in the
hydrophobic domains are glycine, valine, alanine, and proline. The hydrophilic domains are mainly composed of lysine interspersed by alanine. The insoluble elastic fiber is formed via the enzymatic crosslinking of tropoelastin. To facilitate the crosslinking, tropoelastins must associate and align via the process of coacervation triggered by increasing temperatures. The tropoelastin molecules are then enzymatically crosslinked via the lysine residues in the hydrophilic domains following the self-assembly process and eventually yield desmosine and isodesmosine linkages. The interchain crosslinking and the high number of hydrophobic residues immediately lead to insolubility of the elastin. Despite the intensive studies from many research groups, details of the structure-function relationship that gives elastin both the ability to self-assemble and its elastomeric properties are still a subject of investigation and debate. However, the hydrophobic domains have been suggested to be responsible both for the self-assembly properties of tropoelastin as well as for its ability to act as an elastomer.

A new class of synthetic polypeptides with elastin-based sequences has been developed and is known as elastin-like polypeptides (ELP). The model polymers consisting of a pseudo-periodic, low complexity sequence with several types of repeat motifs have been investigated by several groups to obtain a better understanding of the origin of the interesting characteristics of elastin and its precursor. Urry performed pioneering work in the development of ELP by chemically synthesizing poly(VPGVG). VPGVG sequence, which was known to recur in elastin and tropoelastin, has been widely used as a model of ELP. This pentapeptide motif was also generalized into the VPGXG by Urry’s group, where fourth X is often termed the “guest” residue, as it can be substituted with other amino acids except for
proline. It was observed that poly(VPGXG) polymers coacervate in a similar way to tropoelastin and therefore exhibit the LCST behavior or inverse transition temperature ($T_t$) behavior. Substituting the X residue provides a precise molecular parameter to tune the $T_t$. Other pentamer motifs with variation from the VPGXG motif such as LPGXG, IPGXG and VPAVG, and hexamer polypeptide such as VGVAPG, VPGVAG, AVPGVG, VPAGVG, and VPGVAG have also been extensively used. Chilkoti and coworkers synthesized ELP with nine different alanine insertion and substitution mutants of poly(VPGVG) and then characterized the thermally responsive behavior of this family of ELPs. It was demonstrated that the entire set of ELPs exhibit thermally responsive behaviors, which can be easily and quantitatively tuned by chain length and concentration of solute. Moreover, despite having the same amino acid composition and molecular weight, the $T_t$s of polypeptides composed of repeats of AVPGVG and VPAGVG, which only differ in amino acid arrangement, are substantially different, indicating that the overall hydrophobicity of these polypeptides fails to explain their propensity for coacervation. This work should motivate the intensive study of the LCST behavior of ELP which can provide more detailed level understanding.

1.2.2 Deoxyribonucleic acid (DNA)

Natural DNA is composed of four types of nucleosides: adenine (A), guanine (G), thymine (T), and cytosine (C). These bases can recognize each other through preferential Watson-Crick base pairing of G to C and A to U. DNA, commonly known as the molecule responsible for storage of genetic information in the cell, is a rising star in the technological
arena mainly due to its inherent molecular recognition ability. From a materials science perspective, DNA is a programmable polymer whose physico-chemical properties depend on its nucleotide sequence. The Watson-Crick base pairing can be employed to obtain the directed functionalization of modified single-stranded DNA.

The use of DNA in nanotechnology enables exploration of DNA’s biological, structural, and functional properties in unprecedented detail\(^2^5\). The construction of DNA-based nanomaterials is mainly based on the process of self-assembly, such as the hybridization of two complementary single-stranded DNA sequences into double-helical strands. Moreover, hybridization of two single DNA strands (ssDNA) into a double helix (dsDNA) leads to changes in both its mechanical and electrical properties. DNA is a highly charged biopolymer with a charge density of one negative charge per base which means that its structure depends significantly on the properties of the solvent in which it is immersed. The ability to recognize base-pair complements and the accompanying change in materials properties are what drive the use of DNA in novel materials\(^2^6\) such as DNA-based biosensors,\(^2^6\)\(^-^2^7\) DNA-functionalized colloidal materials,\(^2^8\) DNA templates for biomineralization,\(^2^9\) 2D and 3D DNA origami,\(^3^0\) and DNA-based directed surface assembly\(^3^1\). Overall, DNA-based materials have the ability to alter the future of materials science, nanotechnology and nanobiology, but achieving this goal will require obtaining systematic molecular insights into the fundamental aspects of DNA dynamics and structures and the interplay between DNA and nanostructures.\(^3^2\) Moreover, investigation of DNAs properties in
nanotechnological applications will allow us to build a more complete picture of the molecular biology of natural DNA.

1.2.3 DNA block copolymers

DNA has also been used to create hybrid amphiphiles\textsuperscript{33} composed of hydrophilic oligonucleotides and synthetic, hydrophobic polymers.\textsuperscript{34} Since DNA represents a polyelectrolyte, DNA block copolymers, which combine structural features of polyelectrolyte, block copolymers and surfactants, are expected to show a rich association behavior.\textsuperscript{35} DNA block copolymers have been widely exploited for the delivery of antisense DNA\textsuperscript{12} and SiRNA \textsuperscript{36} as well as for the fabrication of self-assembled materials such as multiblock copolymers, micelles and vesicles\textsuperscript{14,34c,d,37}.

Several template-based synthesis strategies, such as polymerase chain reaction (PCR),\textsuperscript{37a} enzymatic ligation\textsuperscript{38} and rolling circle amplification\textsuperscript{39} have been employed to synthesize single-strand DNA. Most recently, Tang et al. reported on a simple, biomimetic synthesis approach that harnesses an enzyme-catalyzed polymerization reaction to directly synthesize high MW, single-stranded DNA (ssDNA) amphiphiles in solution and with low polydispersity. This enzyme-catalyzed polymerization of DNA is conceptually similar to controlled or “living” polymerization of synthetic polymers, and opens up new directions of research and potentially new applications for this new class of DNA block copolymers. With the development of new and sophisticated synthesis techniques, the design of a variety of DNA block copolymers with a broad range of properties and characteristics becomes feasible.
1.3 Methods

1.3.1 Molecular Dynamics

Molecular dynamics (MD) is an important tool for studying the structure, thermodynamics, and interactions of biomolecules because of its unique ability to simulate these systems at atomic resolution over time scales approaching those of biological processes. The combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates is commonly referred to as a “force field”. The force field functional forms can be classified into two groups: bonded and non-bonded. The total interaction energy is

\[ E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} + E_{\text{other}} \]  

Where \( E_{\text{bonded}} \) is the contribution to the total energy from bonded interactions and seeks to account for the stretching of bond, the bending of valence angles, and the rotation of dihedrals; \( E_{\text{nonbonded}} \) is the contribution from nonbonded interactions and includes electrostatics, dispersions, and Pauli exclusions; \( E_{\text{other}} \) includes any force field-specific terms. The quality of molecular dynamics (MD) simulations relies heavily on the accuracy of the underlying force field. Considerable time and effort has been spent on increasing the accuracy of MD force fields to increase both the predictive and analytic capabilities of MD simulations. The currently available all atom force fields have significantly evolved since their inception in the 1980s. Several protein force fields have been widely used, including CHARMM\textsuperscript{55}, AMBER\textsuperscript{56}, GORMOS\textsuperscript{57}, and OPLS-AA\textsuperscript{58}. 

9
The functional form of the Amber force field is

\[ E = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2 + \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \]

<table>
<thead>
<tr>
<th>Bonds</th>
<th>Angles</th>
<th>Torsions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sum_{j=1}^{N} \sum_{i=j+1}^{N} f_{ij} \left{ \epsilon_{ij} \left( \frac{r_{ij}}{r_{ij}} \right)^{12} - 2 \frac{r_{ij}}{r_{ij}} \right} )</td>
<td></td>
<td>(\sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)])</td>
</tr>
</tbody>
</table>

Systematic and extensive evaluation of different protein force fields based on comparisons of experimental data with MD simulations has been obtained in several works.\(^{59}\)

It was suggested that the most recent versions of protein force field, while not perfect, provide an accurate description of many structural and dynamical properties of proteins. Modifications of exiting force fields have led to improvement in agreement with experimental data. However, certain controversy exists with the comparison of secondary structures of the protein using different force fields, which is due to possible biases different force forces have toward certain type of secondary structures.\(^{60}\) Capturing the cooperativity of helix and hairpin formation and melting (denaturation) appears to be a general area for further improvement of force fields.

Atomistic simulations of protein and protein-based systems have been performed using both explicit representations of water and implicit generalized Born models. A number of explicit water models are developed in order to describe different target properties, such as
The choice of water model strongly influences the accuracy of the calculated physical properties. SPC and SPC/E are polarized models and the SPC/E model is a modified SPC model, taking into account the self-energy correction. The SPC/E model showed a better performance when presenting density and diffusion constant than the SPC model. In TIP4P model, a dummy site without mass is introduced where the negative charge is placed. Previous studies showed that only TIP4P provides a qualitatively correct phase diagram on water among TIP3P, SPC, SPC/E, TIP4P and TIP5P. TIP4P model provides better values for the density maximum, temperature and expansion coefficient. However, due to the dummy site, the usage of TIP4P model induces high computation cost. The performance of water models is quite different when they are applied to determine a particular physical property. Generally, The SPC/E water model performs best when simulating the aqueous biomolecular systems. In EXP6 model, a switch from Lennard-Jones potential to Buckingham exponential-6 potential to describe the nonpolar interactions is made. EXP6 model reproduced accurately the vapor pressure of water, however, poor description of the dielectric constant. So far, no water model available is able to reproduce all the water properties with good accuracy. It is worth considering which models to use when performing any simulation including water.

There are several algorithm to solve the Newton’s Equations of motion: Verlet algorithm, Leap Frog algorithm, and Velocity Verlet algorithm. The most common time integration method is the Verlet algorithm, which uses positions and accelerations at time $t$ and the position from time $t - \Delta t$ to calculate new positions at the time $t + \Delta t$. 
\[ r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t) \Delta t^2 + O(\Delta t^4) \] (1.3)

The time step $\Delta t$ must be very small to capture the fastest motions in a system which
is commonly chosen to be on the order of 1fs. In molecular dynamics, successive
configurations can be generated by the simulation. Based on Ergodic Hypothesis, that an
ensemble average (which relates to many replicas of the system) is the same as an average
over time of one replica, an ensemble average of thermodynamic variables of the simulated
system can be obtained with the configurations generated from MD simulation when the
system reaches equilibration.

1.3.2 Dissipative Particle Dynamics

Dissipative Particle Dynamics (DPD), a meso-scale simulation technique, is used to
study and predict the phase behavior and properties of copolymer aggregates as well as
complex, self-assembled structures such as multicomponent micelles and polymersomes.\(^{49}\) In
DPD, a number of molecular entities are coarse-grained into an element, thereafter called a
DPD bead.\(^{49b, 50}\) These DPD beads are subject to soft potentials and governed by predefined
collision rules. The beads move according to Newton’s equations of motion. For a DPD bead
i, we have

\[
\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i , \quad m_i \frac{d\mathbf{v}_i}{dt} = \sum_{j \neq i} \mathbf{f}_{ij} , \quad \mathbf{f}_{ij} = (\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R) + \mathbf{F}_{ij}^S \quad (1.4)
\]

where $m_i$, $\mathbf{r}_i$, and $\mathbf{v}_i$ are the mass, position, and velocity of bead i, respectively. $\mathbf{f}_{ij}$ is
the interbead force on bead i by bead j. $\mathbf{F}_{ij}^C$, $\mathbf{F}_{ij}^D$, and $\mathbf{F}_{ij}^R$ are the conservative force,
dissipative force, and the random force, respectively, acting between beads $i$ and $j$. They are given by

\[
F_{ij}^C = \begin{cases} 
  a_{ij} (1 - r_{ij} / r_c) \mathbf{r}_{ij}, & r_{ij} < r_c \\
  0 & r_{ij} \geq r_c
\end{cases}
\]

\[
F_{ij}^D = -\gamma w^D(r_{ij}) (\mathbf{r}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{r}_{ij}
\]

\[
F_{ij}^R = \sigma_D w^R(r_{ij}) \theta_y \Delta t \frac{1}{2} \mathbf{r}_{ij}
\]

(1.5)

where $a_{ij}$ is the maximum repulsion between beads $i$ and $j$; $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $r_{ij} = |\mathbf{r}_{ij}|$, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$, and $\mathbf{r}_{ij} = \mathbf{r}_j / r_{ij}$ is the unit vector directed along $j$ to $i$, $\gamma$ and $\sigma_D$ are coefficients characterizing the strengths of the dissipative and random forces, $\Delta t$ is the iteration time step, and $\theta_y(t)$ is a symmetric random variable.

Español and Warren\textsuperscript{51} showed that for the fluctuation-dissipation theory to be satisfied at some temperature $T$, the two weight functions and their prefactors must satisfy the following relationship:

\[
w^D(r) = [w^R(r)]^2, \quad \gamma = \frac{\sigma_D^2}{2k_B T}, \quad w^D(r) = [w^R(r)]^2 = \begin{cases} 
  (1 - r / r_c)^2, & r < r_c \\
  0 & r \geq r_c
\end{cases}
\]

(1.6)

One of the major advantage of DPD is the intuitiveness and ease with which simple models for various complex fluids can be constructed by modifying the conservative interactions (repulsive parameter $a_{ij}$) between DPD beads, although there is certain limitation for DPD fluid to exhibit a rigid thermodynamic behavior of the system.\textsuperscript{52} However, it was demonstrated that the coarse-grained mesoscopic model correctly reproduces the properties of a real system beyond certain length and time scales.
Conventional DPD, where the repulsive parameters for each component $a_{\text{polymer-polymer}} = a_{\text{solvent-solvent}}$, can successfully predict the phase behavior of monomer-monomer and polymer-solvent mixtures.$^{53}$

1.4 LCST behavior of polymers and biopolymers

Polymers exhibiting LCST behaviors include a class “smart” polymers that are highly sensitive to temperature and respond with a sharp change in their solubility. The LCST behavior is usually detected by turbidity measurement, which precisely describes the macroscopic transition. It was proposed that there are generally three types of polymers exhibiting LCST behavior.$^1$ For polymers in type I, the LCST shifts upon increasing the polymer molar mass towards lower polymer concentration. The LCST of polymers of type II is hardly affected by chain length. The architecture also has a negligible effect. Type III polymers exhibit a bimodal phase diagram, presenting two critical points at low and at high polymer concentrations. The LCST behavior of polymers is usually considered to be regarded as the consequence of the competition between polymer-water interactions and polymer-polymer interactions. At low temperatures, the polymer-water interactions are favorable which allow solubilization, while above LCST hydrophobic polymer-polymer interactions are thermodynamically favored which result in the self-aggregation of polymer chains. The LCST was first described by Heskins and Guillet for an aqueous solution of PiPAAm.$^{41}$ It was evident that the formation of nonpolar and intermolecular hydrogen bonding is essential to the LCST behavior.$^{41}$ The most widely studied water-soluble LCST
polymer, poly(N-isopropylacrylamide), which is exhibited to undergo a coil-globule 
transition and form more intrachain hydrogen bonds upon heating above its LCST. However, 
for some polymer chains without hydrogen bond donors, such as POEGMA and PDEAM, 
are not able to form any intra- or inter-chain hydrogen, the phase transition is mainly arising 
from the multiple chain aggregation without a precontraction process of individual polymer 
chains.42

ELPs are inherently biodegradable and biocompatible with higher sensitivity to salt as 
compared to carbon-backbone polymers such as poly(N-isopropylacrylamide) (PNIPAM).43 
As ELPs are biodegradable and biocompatible, ELPs may probably become the mostly 
applied synthetic temperature-sensitive biopolymers in biotechnology applications.

ELPs that exhibit LCST behavior may also display temperature-dependent hysteresis 
in their phase transition behavior. For example, it was demonstrated that poly(VPAVG) 
display significant thermal hysteresis upon cooling.23c The hysteretic phase behavior of ELPs 
is poorly understood and largely unexplored. In a recent study from Chilkoti’s group, the 
thermo hysteresis found for poly(VPAGVG) showed similarities but also significant 
differences to poly(AVPGVG).11c In a word, the aggregation and disaggregation processes 
for ELP must be considered as complex multistep process that cannot be satisfactorily 
described by conventional kinetic models and analysis.
1.5 Self-assembly of amiphiphiles

From the thermodynamics point of view, the structure obtained through self-assembly of amphiphilic block copolymers is dominated by the free energy arising from the (electrostatic or/and steric) repulsion between the coronal blocks and the excess free energy of the core-water interface. Amphiphilic block copolymers may self-assemble in a variety of nanoaggregates in aqueous environment, such as micelles, vesicles and lamellar mesophases or micellar aggregates. Among these morphologies, the three most common ones are spherical micelles (including star-like and crew-cut), cylindrical micelles, and vesicles. The morphology and size of aggregates formed is determined by the characteristically complex equilibrium of noncovalent forces (electrostatic, steric, hydrogen bonding, Van der Waals, and hydrophobic interactions) and can be efficiently tuned by variations hydrophobic/hydrophilic chain length, solvent ionic strength or/and pH in the aqueous solution.

The influence of salt on the micellization properties was also presented for polyelectrolyte copolymers which combine structural features of polyelectrolyte, block copolymers and surfactants. It was demonstrated that the strength of repulsive Coulomb interactions between the polyelectrolyte (PE) segments can be screened by the addition of salt if solvent ionic strength is higher than the concentration of counterions and thus the solvent ionic strength plays a decisive role in the equilibrium properties in these systems. The addition of salt may induce shrinking of corona chains, increases in aggregation number and the shape transition.
Chapter 2 provides a review of the recent progress in molecular modeling of DNA materials. In Chapter 3 and 4, a molecular description of the LCST behaviors of Elastin-like polypeptide (VPGVG)$_n$ is provided. In Chapter 5 and 6, the role of amino acid sequence on the LCST behaviors of ELPs is described. Chapter 7 introduces an Implicit Solvent Ionic Strength (ISIS) method to model polyelectrolyte systems with Dissipative Particle Dynamics (DPD). In Chapter 8, 9 and 10, the effect of solvent and block properties on the self-assembly of polyelectrolyte block copolymers is investigated.
1.6 References


18. Sandberg, L. B.; Weissman, N.; Smith, D. W., Purification and Partial Characterization of a Soluble Elastin-Like Protein from Copper-Deficient Porcine Aorta. Biochemistry 1969, 8 (7), 2940-


the mixture of diblock copolymer poly(ethyl ethylene)-block-poly(ethylene oxide) and homopolymer poly(propylene oxide) in aqueous solution. *Polymer* 2009, *50* (22), 5333-5340.


92. (a) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J., A Mesoscale Model of DNA and Its Renaturation. *Biophys J* 2009, 96 (5), 1675-1690; (b) Knotts, T. A.; Rathore, N.; Schwartz,
D. C.; de Pablo, J. J., A coarse grain model for DNA. *The Journal of Chemical Physics* 2007, *126* (8), -.


Chapter 2 Progress in molecular modeling of DNA materials

Nan K. Li, Ho shin Kim, Jessica A. Nash, Mina Lim, Yaroslava G. Yingling

Published in Molecular Simulation 40 (10-11), 777-783

The unique molecular recognition properties of DNA molecule, which store genetic information in cells, are responsible for the rise of DNA nanotechnology. In this paper, we review recent advances in atomistic and coarse-grained force fields along with simulations of DNA-based materials, as applied to DNA-nanoparticle assemblies for controlled material morphology, DNA-surface interactions for biosensor development and DNA-origami. Evidently, currently available atomistic and coarse-grained representations of DNA are now at the stage of successfully reproducing and explaining experimentally observed phenomena. However, there is a clear need for the development of atomistic force fields which are robust at long time scales and in the improvement of the coarse grained models.

2.1 Introduction

The use of DNA in nanotechnology enables exploration of DNA’s biological, structural, and functional properties in unprecedented detail. DNA, commonly known as the molecule responsible for storage of genetic information in the cell, is a rising star in the technological arena mainly due to its inherent molecular recognition ability. From a materials science perspective, DNA is a programmable polymer whose physico-chemical properties
depend on its nucleotide sequence. The construction of DNA-based nanomaterials is mainly based on the process of self-assembly, such as the hybridization of two complementary single-stranded DNA sequences into double-helical strands. Natural DNA is a string of four types of nucleosides: adenine (A), guanine (G), thymine (T), and cytosine (C). These bases can recognize each other through preferential Watson-Crick base pairing of G to C and A to U. Moreover, hybridization of two single DNA strands (ssDNA) into a double helix (dsDNA) leads to changes in both its mechanical and electrical properties. The DNA persistence length, which is an indicator of the chain’s flexibility, changes from ~1nm for ssDNA to ~50nm for dsDNA depending upon sequence and solvent conditions. Therefore, the mechanical properties of a 100-mer ssDNA, which is highly flexible, will be significantly different from that of a 100-mer dsDNA, which is relatively rigid at this length. Hybridization also changes the electrical properties of DNA since dsDNA has considerably greater charge transport properties than ssDNA. DNA is a highly charged biopolymer with a charge density of one negative charge per base which means that its structure depends significantly on the properties of the solvent in which it is immersed. The ability to recognize base-pair complements and the accompanying change in materials properties are what drive the use of DNA in novel materials such as DNA-based biosensors, DNA-functionalized colloidal materials, DNA templates for biomineralization, 2D and 3D DNA origami, and DNA-based directed surface assembly. Overall, DNA-based materials have the ability to alter the future of materials science, nanotechnology and nanobiology, but achieving this goal will require obtaining systematic molecular insights into the fundamental aspects of DNA
dynamics and structures and the interplay between DNA and nanostructures. Moreover, investigation of DNAs properties in nanotechnological applications will allow us to build a more complete picture of the molecular biology of natural DNA.

Since experimental methods for detailed studying of dynamics and structure of DNA and DNA-functionalized materials are yet to be developed, it is useful to employ a complimentary approach that allows for direct visualization of the processes and interactions involved. Computational techniques have proven to be essential in elucidating dynamical processes in biomaterials, DNA, nanomaterials and peptide-surface interactions. Processes involved in the design of the DNA-based nanoscale materials have been addressed by a wide variety of theoretical and computational techniques, including analytical theories, density functional theory, molecular and Langevin dynamics simulations, coarse-grained models and Monte Carlo calculations. Analytical theories can be used to predict the structure of DNA in solution or hybridization kinetics and thermodynamics. Recently developed novel coarse-grained models can be used to predict the crystallization of DNA-functionalized nanoparticles.

Substantial advances in computational resources (faster processes, GPU-enabled simulations, infiniband, etc.) and the development of novel algorithms (implicit solvent, self-guided Langevin dynamics, replica exchange simulations, new parallelization algorithms, etc.) enable molecular simulations to address larger systems and tackle the critical issues in the area of DNA-based materials. Herein, we will discuss only a few key topics, as there are a great number of excellent review articles that can be used to explore the rich filed of
simulations and modeling study of DNA-based materials in greater depth.\textsuperscript{13} We will first review the advances in molecular modeling force fields and representations and then review the recent progress in modeling of DNA-based materials on atomistic and coarse-grained molecular scale.

\textbf{2.2 Molecular Modeling}

\textit{2.2.1 Atomistic simulation}

Atomistic molecular dynamics simulations can provide a complete microscopic description of the structure and dynamics of DNA under different environmental conditions, from detailed information on atom-to-atom interactions via hydrogen bonding or pi-pi stacking to atom interactions with ions, small molecules, and proteins to global functionally important motions and conformational changes which control the processes of self-assembly. Using an accurate force field and efficient algorithm detailed all-atom investigation can shed the light on the effect of various mono and divalent ionic concentrations, temperature and protein binding on the structural properties of key DNA building blocks that participate in self-assembly, such as single strands, helices, loops, bulges, loop-loop motifs and junctions\textsuperscript{13a, 13d}.

Atomistic molecular modeling represents one promising avenue for the rational prediction and engineering of DNA-based systems. Atomistic simulation studies have been conducted to examine the complexity of, for example, DNA-carbon nanoparticle/graphene\textsuperscript{14}, complexity of DNA-Au surface/nanoparticle \textsuperscript{15}, and the binding of DNA-protein\textsuperscript{16}. The atomistic simulation protocols and force fields for DNA-related systems are mature and
robust. Significant efforts are also devoted to the optimization of performance using hardware accelerators and to the development of the advanced simulation and analysis workflows. Currently, using advanced computational resources, the atomistic simulations of very large systems, such as DNA-NPs system consisting of millions of atoms, or very long simulations, such as the microsecond simulations of B-DNA helical dynamics, are now a reality.

The results of the atomistic simulations are dependent on the accuracy of the force field. Several different all-atom force fields have been developed for DNA, including CHARMM, AMBER, Bristol-Myers Squibb (BMS), and GORMOS. The most commonly used force fields for DNA studies are AMBER and CHARMM. However, many efforts have been devoted in the constant refinement of current empirical force fields for DNA. For example, a problem was identified in simulations of longer than 10 ns which involved treatment of the α/γ torsions in the Amber Cornell et al. force fields. The applied bsc0 correction on the basis of quantum mechanical (QM) calculation on model representative of the phosphodiester backbone stabilizes simulations of DNA duplexes on the microsecond time scale.

There are efforts to develop new force fields for specific DNA systems. Force field parameters for LNA residues (locked nucleic acid) were created by parameterization of glycosidic χ angles and RESP charges on the basis of QM calculations. As the result of this development, an improved agreement between computational predictions and NMR spectra was achieved. Another exciting progress has been in the development of polarizable DNA
force fields. For example, Baker et al. developed a polarizable force field for nucleic acid bases based on the CHARMM Drude polarizable force field. An improvement in the reproduction of heats of sublimation over a non-polarizable force field was reported. All force fields are always under development as larger ensembles and longer trajectories reveal previously unobserved deficiencies. Even minor inaccuracies may have major effects on populations of different structural forms, leading to artificial geometries and instable trajectories at long time scales.

2.2.2 Coarse-grained model

Most of interesting biological and material science problems, which typically involve DNA-based materials, occur on the relatively large system sizes and long time scales that are currently unattainable with all-atom simulations. For instance, all-atom models are simply too computationally expensive to allow studies on the time scales associated with self-assembly. Thus, there is a need for the development of coarse-grained models which are based on the “integration” of large number of degrees of freedom into a few coarse-grained particles or beads to achieve a good balance between reasonable physical description and computational feasibility. In general, as the number of beads per nucleic base decreases, the simulation become less computationally expensive and the system size and time can be increased. However, building a force-field that is both accurate and transferable becomes more difficult as the graining becomes “coarser”, because more specific interactions must effectively be included with fewer parameters and functional forms.
In a highly coarse grained models of DNA, either single strand or double stranded DNA are to be considered as a chains of spheres that are connected by springs. The bead-and-spring models based on Brownian Dynamics\textsuperscript{31} or Dissipative Particle Dynamics\textsuperscript{32} provides excellent agreement with experimental data for a variety of rheological properties or self-assembly behaviors in aqueous solutions\textsuperscript{33}, such as the micellization of DNA amphiphiles.\textsuperscript{34} The simplified model has also been shown to reproduce the behavior of small systems of 6-8 base pairs near a surface.\textsuperscript{35}

However, for some applications of interest, the aforementioned bead-and-spring models are inadequate to provide the resolution or molecular detail required to describe a variety of phenomena, such as hybridization, binding of proteins, and melting. Further refined coarse-grained potentials have been developed to capture the key effects in the relevant phenomena occurring at a much more localized level. By introducing a Morse potential for hydrogen bonding interactions and harmonic potentials for backbone maintenance, a one bead-per-base model of double strand DNA by Doi et al.\textsuperscript{36} was able to reproduce salt dependence of a persistence length and melting behavior.

Several three-bead models, i.e. three sites per nucleotide, have been proposed with improved specificity of local interactions. De Pablo\textsuperscript{37} developed a three bead-model with a parameterization optimized by fitting thermal properties from all-atom simulations of short DNA chains. In the three bead mapping, each nucleotide is represented by three beads corresponding to phosphate, sugar and base functional groups. Additional water-ion potential was incorporated in the three-bead model that allows studying the features of solvation
Another three-bead CG model targeted to reproduce the thermodynamics of DNA melting was presented to successfully capture the essential structural properties of single-strand and double strand DNA, which was applied to the study of DNA nanotwizers. At last, a six-bead model was developed by Pantano and coworkers. The nucleotide is represented by six beads, whereas the chemical skeleton of the all-atom nucleotide is presented explicitly. This model has been applied to the study of ion-DNA binding, the A->B structural transformations, and the thermal melting, which is proven to be very efficient in predicting the structural details.

2.3 Modeling DNA-based materials

2.3.1 DNA-nanoparticle assemblies

There are many studies that show how a controlled DNA hybridization process can be used to facilitate the programmed assembly of the materials components with nanoscale resolution. Functionalization of nano and micro-scale objects with DNA leads to programmable materials assembly into 2D and 3D networks and geometries. In 1996, Mirkin et al and Alivisatos et al showed that gold nanoparticles could be assembled using DNA. Gold nanoparticles (AuNP) of the same or different sizes were reacted with thiol-modified ssDNA and hybridization between complementary ssDNA led to the formation of a 3D network. The ability of the DNA-AuNP network to form a super crystal depended on DNA length, sequence, DNA mechanical properties (ssDNA versus dsDNA), AuNP size and...
polydispersity. Moreover, dense packing of dsDNA onto 5 and 10 nm gold nanoparticles can induce length-dependent denaturation of DNA at high surface density.\textsuperscript{45}

Both atomistic and coarse grained simulations have been used to study properties of crystals made from DNA functionalized nanoparticles\textsuperscript{12a,12c, 19, 46}. Ngo et al\textsuperscript{19} used atomistic molecular dynamics simulations of solvated nanoparticle/DNA lattices to compare body-centered cubic and face-centered cubic supercrystals in systems containing millions of atoms. Atomistic simulation allowed for the observation of changes in the canonical B-DNA structure, ion distribution in the lattices and ion binding energy. Properties such as Young’s and bulk moduli were also measured. Results suggested that lattices did not preserve DNA in its B-structure. Although an atomistic approach gives valuable insight into the atomistic properties of the system, the self-assembly properties of the structures cannot be studied. Self-assembly processes take place on a longer time scale than is practical in atomistic simulations. In order to study crystallization of DNA-NP lattices, coarse grained methods are employed. In coarse-grained simulations, results are qualitative rather than quantitative - system details such as DNA helical conformation and pi-pi stacking may not be measured. Li and coworkers\textsuperscript{46} were able to simulate assembly of different lattice structures using a coarse-grained model by varying nanoparticle stoichiometry, size and DNA grafting density. The formation of BCC, CsCl, AlB\textsubscript{2} or Cr\textsubscript{3}Si crystal structures was shown to depend on particle size, particle stoichiometry or DNA strand length.
2.3.2 DNA functionalized on gold surface

Surfaces functionalized with DNA can be used for biosensing moieties\(^7\).\(^{41}\), where surface-immobilized ssDNA binds target ssDNA in solution with the hybridization event monitored by strategies including labeling with fluorescent tags, detecting binding of reporter molecules to dsDNA, or detecting changes in materials properties (mechanical, electrical).\(^{3a}\) Moreover, DNA has a sequence-specific binding affinity to gold surfaces and the strength of such interactions depends on the DNA surface density and length, solvent, thiol linker and presence of double helix.\(^{47}\) The interplay between the surface of nano and micro-scale objects, DNA length and sequence, linker design and self-assembly via hybridization is not clear and further progress in the development of DNA-based materials highly depends on understanding of the structures and dynamics of DNA in a precisely way. Fundamental understanding of dynamical DNA adsorption behavior is required for the advances in the future development of biosensors and optimization of DNA microarray\(^{48}\).

All atom MD simulation has been spotlighted since the determination of DNA structure and dynamics on surfaces can be most accurately accomplished using all-atom models\(^{49}\). Wong et al. performed the first all-atom MD simulation of DNA grafted on silica surfaces to elucidate the structure of DNA and their interactions on microarray surface\(^{15f}\). This early study showed that binding between DNAs was enhanced by tilting towards the nearest neighbor and forming close packed conformations. Lee et al. carried out atomistic MD simulations for a DNA functionalized gold surface\(^{50}\). In this study, it was found that base stacking and non-Watson-Crick hydrogen bonding lead to close contact between dangling
end of the DNA and its counterpart on neighboring DNAs at specific DNA density on gold surface. This study concluded that these DNA-DNA interactions play a pivotal role in stabilizing closely packed aggregates of DNA-functionalized nanoparticles. In the other study by the same group the differences between a gold surface with a low coverage and that with a high DNA coverage was examined via MD simulation as well as the processes on the surface of small gold nanoparticle. They showed that low coverage of gold surface results in conformational stabilization of ds-DNA as a Watson-Crick B-DNA and obtained the effective radius of DNA-gold nanoparticle. Also, it was investigated that increase in sodium concentration causes the increase in melting temperature DNA, which agrees with the experiment regarding DNA-linked gold nanoparticle aggregates. Our group used all atom MD simulations in order to establish the effect of surface immobilization on DNA dynamics. We found that the possibility of cross-strand interactions is dependent on the length and sequence of single stranded DNA. To investigate the mechanical behavior of DNA tethered on gold surface and its dynamics, Wang et al. performed non-equilibrium all-atom MD simulation in nanochannel shear flows. By comparing tethered DNA with free DNA, effect of shear stress, differences based on conformation, and hydrodynamic forces were determined. In the light of aforementioned results, the author suggested that the results can be of use in extending macroscopic empirical models of DNA dynamics in shear flows.

2.3.3 Graphene and DNA

Graphene possess different affinities for single stranded versus double stranded DNA, which makes it an ideal material for novel high efficient sensor development. Moreover,
ssDNA can form various morphologies on graphene. For example, Akca et al.\textsuperscript{53} found that when ssDNA is added to graphene solution two competing \(\pi-\pi\) stacking interactions can take place, which are between DNA bases and between DNA base and graphene. The competition between these two stacking events leads to formation of two distinct DNA assembly patterns on graphene, which are small spherical particles or elongated networks. However, simulations of interactions of ssDNA with graphene are rarely discussed as compared to carbon nanotubes.

Recent atomistic molecular dynamics simulations were used to examine the interactions of dsDNA and ssDNA with graphene. Zhao\textsuperscript{14b} performed molecular simulations to explore various self-assembled structures that dsDNA can form on graphene surfaces and nanotube arrays. They show that \(\pi-\pi\) interactions are the dominant force in the self-assembly mechanism between DNA and carbon-based materials, which even allows for dsDNA segments to stand up on the surface. Furthermore, Chan et al\textsuperscript{54} shown that dsDNA can be trapped between graphene layers in standing up conformation. Recently, Manna et al\textsuperscript{14a} used atomistic molecular dynamics simulations to explore ssDNA structure and dynamics on graphene. They found that there are three simultaneous competing forces, which are the \(\pi-\pi\) stacking interactions within ssDNA, H-bonding interactions with solvent water molecules, and the base-graphene stacking interactions, which facilitate ssDNA adsorption on graphene. The relative importance of these interactions lead to the observed differences in the binding free energy, such as \(\text{ssd(AGTC)}_3 < \text{ssdA}_{12} < \text{ssdG}_{12} < \text{ssdC}_{12} < \text{ssdT}_{12}\), which indicated that hetero-sequences are the best for graphene dispersion in aqueous solutions.\textsuperscript{14a}
2.3.4 *Nonspecific DNA and nanoparticle assemblies*

Many DNA superstructures, including DNA origami nanostructures\(^{55}\), and discrete 2D/3D architectures\(^{56, 57}\) have been used to assemble various nanoscale objects such as Au nanoparticles\(^{58, 59}\), quantum dots\(^{60}\), carbon nanotubes\(^{61}\), graphene\(^{62}\), conductive-polymer\(^{63}\), or proteins\(^{64}\), with the potential applications in the fields of electronics, photonics or bionics. DNA’s identity as a polyelectrolyte may be harnessed to create DNA composites with cationic nanoparticles\(^{65}\) or dendrimers\(^{66}\) for applications such as drug delivery or agents for gene transfection.

Molecular dynamics simulations have shown that DNA may wrap around charged cationic dendrimers\(^{67}\) and nanoparticles\(^{68}\), even for short DNA sequences below its persistence length. Interaction of DNA with cationic particles can be controlled by varying nanoparticle size or salt concentration. Zinchenko et al\(^{68}\) used experimental techniques coupled with coarse grained molecular dynamics simulations to identify three modes of interaction of cationic nanoparticles and long strands of DNA, which were represented by a generic polyelectrolyte in the simulations. Depending on nanoparticle size, adsorption, wrapping, or collection occurred for large, medium and small nanoparticles respectively. A combined experimental and all-atom molecular dynamics study of weakly charged nanoparticle adsorption on DNA showed that weakly charged nanoparticles adsorb to DNA without a significant distortion of DNA helical structure. However, very high concentrations of weakly charged cationic nanoparticles have been shown to cause strand separation both experimentally and in molecular dynamics simulations\(^{69}\).
2.3.5 DNA origami

DNA can be folded into arbitrary two and three-dimensional shapes via DNA origami methods\textsuperscript{70}. Scaffold strands of up to several thousand bases may be folded to achieve structures with nanometer-scale dimensions\textsuperscript{70-71}. Although structures may be observed using experimental techniques such as atomic force microscopy or transmission electron microscopy, less is known about origami solution dynamics and microstructure.

Solution behavior of three different DNA origami structures has been studied using all-atom molecular dynamics\textsuperscript{72}. Two rectangular structures with different lattice structures and one origami with a bended structure were created using caDNAno, a program commonly used to design DNA origamis, and converted to all-atom representations. Molecular dynamics reproduced experimentally known structural properties of the origami. Moreover, simulations allowed for the observation of fine details of the origami structures such as counterion location and distance between DNA helices.

Coarse-grained methods allow for the simulation of larger DNA origami systems or longer time-scales than atomistic methods. However, challenges exist for modelling self-assembly due to the complexity of the self-assembly process and size of DNA origami. Recently, the oxDNA model has been reviewed for simulation of DNA in biotechnology applications. The oxDNA model provides a good description of the basic mechanical properties of DNA, and has been shown to reproduce the structure of DNA tetrahedrons, tiles and origamis made up of more than six thousand bases\textsuperscript{73}. Simulation of self-assembly processes has been less successful due to the large number of events that must take place to
accurately assemble an origami structure. Thus, self-assembly simulation of even small DNA structures is an active area of research.

2.4 Conclusions

In this review, we highlighted recent advances of molecular simulations in the field of DNA-based nanomaterials. The currently available atomistic and coarse-grained representations of DNA are now at the stage of successfully reproducing experimentally observed phenomena on nano- to mesoscopic length scales and timescales. Molecular simulations not only can provide an improved molecular level description of DNA biophysics but also aid in visualizing the mechanisms of self-assembly for DNA nanostructures. It is suggested that molecular simulations can play a significant potential role in rational prediction and design of advanced DNA-based materials. However, there is a clear need in the development of atomistic force fields which are robust at long time scales and in the improvement of the predictability and transferability of coarse-grained approaches. Overall, investigation of DNAs properties in nanotechnological applications will allow us to advance the field of DNA-nanotechnology and build a more complete picture of the molecular biology of DNA.

Acknowledgements

The authors wish to thank the National Science Foundation grant CMMI-1150682 and the NSF's Research Triangle MRSEC (DMR-1121107) for the support.
2.5 References


34. Tang, L.; Tjong, V.; Li, N.; Yingling, Y. G.; Chilkoti, A.; Zauscher, S., Enzymatic Polymerization of High Molecular Weight DNA Amphiphiles That Self-Assemble into Star-Like Micelles. *Advanced Materials* 2014, n/a-n/a.


65. Sokolova, V.; Epple, M., Inorganic nanoparticles as carriers of nucleic acids into cells. Angewandte Chemie (International ed. in English) 2008, 47 (8), 1382-95.


73. Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A.; Romano, F.; Å ule, P.; Matek, C.; Snodin, B. E. K.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; Smith, W. P. J., Coarse-
Chapter 3 Molecular description of the LCST behavior of an elastin-like polypeptide

Nan K. Li, Felipe García Quiroz, Carol K. Hall, Ashutosh Chilkoti, and Yaroslava G. Yingling

Published in Biomacromolecules 15 (10), 3522-3530

ABSTRACT. Elastin-like polypeptides (ELPs) with the repeat sequence of VPGVG are widely used as a model system for investigation of lower critical solution temperature (LCST) transition behavior. In this paper, the effect of temperature on the structure, dynamics and association of (VPGVG)$_{18}$ in aqueous solution is investigated using atomistic molecular dynamics simulations. Our simulations show that as the temperature increases the ELP backbones undergo gradual conformational changes, which are attributed to the formation of more ordered secondary structures such as β-strands. In addition, increasing temperature changes the hydrophobicity of the ELP by exposure of hydrophobic valine-side chains to the solvent and hiding of proline residues. Based on our simulations, we conclude that the transition behavior of (VPGVG)$_{18}$ can be attributed to a combination of thermal
disruption of the water network that surrounds the polypeptide, reduction of solvent accessible surface area of the polypeptide, and increase in its hydrophobicity. Simulations of the association of two \((VPGVG)_{18}\) molecules demonstrated that the observed gradual changes in the structural properties of the single polypeptide chain are enough to cause the aggregation of polypeptides above the LCST. These results lead us to propose that the LCST phase behavior of poly(VPGVG) is a collective phenomenon that originates from the correlated gradual changes in single polypeptide structure and the abrupt change in properties of hydration water around the peptide and is a result of a competition between peptide-peptide and peptide-water interactions. This is a computational study of an important intrinsically disordered peptide system that provides an atomic-level description of structural features and interactions that are relevant in the LCST phase behavior.

### 3.1 Introduction

Elastin is a polymeric extracellular matrix protein that is responsible for the extensibility and elastic recoil exhibited by many vertebrate tissues, including skin, lungs, and larger blood vessels. Despite the great diversity in elastin’s structures, some common features seem to have been preserved in elastin during evolution, such as the presence of significant amounts of glycine, proline and several aliphatic residues.\(^1\) Elastin-like
polypeptides (ELPs) are peptide polymers derived from a portion of the hydrophobic domain of elastin, which is defined by a pseudo-periodic, low complexity sequence with several types of repeat motifs. Most importantly ELPs are known to exhibit inverse temperature or lower critical solution temperature (LCST) phase behavior in aqueous solutions, i.e., they undergo a first-order phase transition into polypeptide-rich and water-rich phases upon heating above the so-called cloud point temperature ($T_{LCST}$)\(^2\). This transition is essentially temperature-triggered coacervation and is reversible\(^3\). The lower critical solution temperature $T_{LCST}$ of ELPs is dependent on pH\(^4\), salt\(^5\), polypeptide hydrophobicity\(^6\), and polypeptide length\(^7\). ELPs are also inherently biodegradable and biocompatible with higher sensitivity to salt than synthetic polymers such as poly(N-isopropylacrylamide) (PNIPAM)\(^8\). The stimulus-responsive character of ELPs has led to their use in a variety of applications, including stimuli-triggered molecular actuators for recombinant protein purification\(^9\), drug delivery\(^10\), and as stimuli-responsive materials for tissue engineering\(^11\). However, despite the proliferation of applications of ELPs in recent years, the origin of their LCST behavior is still a matter of controversy\(^12\). A detailed understanding of the physical underpinnings of this intriguing phenomenon is sorely needed to enable further progress into the development of new elastin-based materials.
ELPs with the canonical (VPGVG)$_n$ sequence seen in animal tropoelastin are widely used as the model system to investigate LCST behavior. Moreover, the LCST behavior of poly(VPGVG) has been experimentally observed using a variety of methods including microscopy$^{3,6a,13}$, differential scanning calorimetry$^{14}$, dielectric relaxation$^{15}$ and different spectroscopic techniques (CD, NMR, FT-IR)$^{16}$. The fourth amino acid valine (Val) is often termed to be the “guest” residue, as other amino acid residues except proline can be substituted for it. The identity of this guest provides a precise molecular parameter to tune the $T_{\text{LCST}}$, as hydrophobic residues depress the $T_{\text{LCST}}$ while polar and charged residues raise the $T_{\text{LCST}}$ as compared to the parent poly(VPGVG)$^{6b}$.

The LCST phase behavior of poly(VPGVG) is a complex and multistep phenomenon, which has been proposed by Urry et al.$^{17}$ to include structural transitions, loss of hydration, expulsion of water molecules and physical association of chains. However, the interplay and relevance of these factors as drivers of the LCST phase behavior of poly(VPGVG) is a matter of controversy, particularly as it relates to the structural changes accompanying the phase transition. Urry et al. carried out extensive studies on synthetic poly(VPGVG) and created a model called the β-spiral model, that involves one type-II β-turn per VPGVG pentamer stabilized by intra- and inter-spiral and inter-turn hydrophobic contacts$^{4,6a,18}$. They proposed
that upon heating \((VPGVG)_n\) undergoes conformational changes from random coil to \(\beta\)-spiral, exposing hydrophobic amino acid side chains to the solvent, which leads to aggregation of ELPs due to hydrophobic interactions\(^{18}\). However, a more recent solid-state NMR study indicated that this polypeptide is in a well hydrated and hence soluble state at temperatures below \(T_{\text{LCST}}\) with relatively well defined type-II \(\beta\)-turns centered around the Pro-Gly pair. Thus Urry’s model says that \((VPGVG)_n\) assumes a random coil structure below the LCST and a \(\beta\)-turn rich conformation (\(\beta\)-spiral) above the \(T_{\text{LCST}}\) but the solid state NMR data says the opposite.\(^{16e, 19}\) The conformational state of an ELP below its transition temperature was also elucidated by CD spectroscopy, which showed a significant amount of type-II \(\beta\)-turns below the LCST\(^{20}\). In addition, a study of \((GVGVP)_6\) by Gross et al\(^{21}\) suggested that the polypeptide could adopt a \(\beta\)-sheet structure instead of a \(\beta\)-spiral in the folded state. A solid-state NMR study by Yao and Hong\(^{22}\) suggested that Urry’s model missed the more populated distorted \(\beta\)-strand structure. Clearly, there is considerable disagreement about the nature and degree of the structural changes and ordering upon heating of ELPs, which provides one of the motivations for this study.

Computer simulations have also been employed to explore the LCST mechanism of poly(VPGVG) in aqueous solution. As detailed observation of phase separation is not yet
possible for these systems, molecular dynamics (MD) simulation studies have been focused on the temperature-induced conformational changes of a single ELP in water. A single ELP was chosen due to current computational limitations. Simulations have been performed on short peptides, GVG(VPGVG)\textsuperscript{23} and GVG(VPGVG)_3\textsuperscript{24}, and a polypentapeptide (VPGVG)\textsubscript{18}\textsuperscript{25}. For the peptide GVG(VPGVG), extended structural conformations dominate at all temperatures and presence of some compact structures is observed at high temperature \textsuperscript{23}. The dominance of the extended conformations may be due to the very short length of this peptide, which would make it difficult to maintain the intrachain interactions required for structural compaction. Interestingly, a study of a peptide that is twice as long, GVG(VPGVG)_3, over a long simulation time challenged the idea of ELP folding upon heating. It was argued that below the transition temperature, ELP chains adopt a rigid conformational state with no resemblance to a random coil; but at high temperatures the peptide is highly flexible with the presence of local ordered structural elements.\textsuperscript{24} In order to determine if the LCST phase behavior is an intrinsic property of the pentameric sequence or is the result of a co-operative effect of many pentamers in a polypeptide, detailed simulations of longer chains are needed. Simulations of a much longer ELP, (VPGVG)\textsubscript{18}, but for a relatively short time scale of 9 ns proposed that at high temperature ELP adopts a compact
structure with distorted β-strands, fluctuating turns, buried hydrophobic residues, and main-chain polar atoms that form hydrogen bonds with water. Unfortunately, the length of this simulation was too short to observe and characterize the dynamic processes of the polypeptide backbone.

This paper presents a comprehensive analysis of the temperature dependence of the secondary structure, torsion angles, hydration and dynamics of (VPGVG)$_{18}$ in explicit solvent by very long atomistic MD simulations. Notably, this paper is the first to examine the effect of temperature on the association between two ELP chains, which is the necessary first step in understanding the molecular determinants of the LCST behavior of polypeptides.

### 3.2 Materials and Methods

#### 3.2.1 Single ELP Simulation

Atomistic MD simulations were performed using Amber 11 and the ff99SB force field for proteins. An initial model of Urry’s β-spiral for (VPGVG)$_{18}$ was solvated in explicit water using the TIP3P water model$^{27}$ (Figure 3.1a). The simulation box of about 5.9 nm × 7.7 nm × 5.0 nm size contained a single peptide and 6,856 water molecules. The size of the box was chosen to be large enough so that no peptide-peptide interactions through periodic boundary conditions can occur, which represent an infinitely dilute state. The simulations
indicated that the initial configuration of the idealized β-spiral model is not stable at all temperatures, which is in agreement with previous simulations\(^{24-25}\) (Figure 3.1). In order to span the temperature range that is relevant to the experimental study of LCST phase transitions in aqueous solutions, simulations were performed at ten temperatures between 290 K and 350 K.

In all cases, each system was equilibrated in eight stages starting from a solvent minimization using steepest descent method for 10,000 steps while keeping the all peptide atoms restrained with 200 kcal/mol. The system was then gradually heated to an assigned temperature for 10 ps while maintaining the 200 kcal/mol constraint on the peptide. A short 40 ps isothermal-isobaric (NPT) ensemble MD run was then performed, again with the peptide restraint maintained at 200 kcal/mol. Another minimization step followed for 10,000 steps with the restraint of 25 kcal/mol. A second NPT MD run was performed at 25 kcal/mol restraint for 20 ps. Subsequently, a final unconstrained minimization of 1,000 cycles was performed before reheating the system to the assigned temperature at constant volume for 40 ps. The NPT ensemble was adopted for the equilibration and production MD run to ensure uniformity in solvent density. The long-range electrostatic interactions were calculated by particle mesh Ewald (PME) summation\(^{28}\) and the non-bonded interactions were truncated at a
9 Å cutoff along with a 0.00001 tolerance for Ewald convergence. The temperature was maintained using a Berendsen thermostat. The SHAKE algorithm was used to constrain the position of the hydrogen bonds. The production simulations were performed for at least 70 ns with a 2 fs time step. Only the last 40 ns of the trajectories from each case were considered for statistical analysis. The conversion of the simulations was assured by the data presented in the Supplemental Information (Figures 3.S3b, 3.S4, 3.S5 and 3.S7). MD trajectories were processed using in-house scripts along with the standard tool suite accompanying Amber11.0. The interaction energy was calculated using the molecular mechanics energy function in NAMD 2.7. The hydrogen bond analysis was performed using a distance cutoff of 3.5 Å and an angle cut-off of 30°.

### 3.2.2 Interactions between two ELPs

In order to get a better understanding of how the temperature-dependent single peptide properties can cause the aggregation between peptides, we simulated the interaction between two temperature-equilibrated peptides. The initial configurations of peptides used in these simulations were chosen through the application to the MD simulation trajectories described above of a hierarchical RMSD-based cluster algorithms following by the energetic analysis (Figure 3.S1, Table S1). Specifically, the last 40 ns of a trajectory from the
single-peptide MD simulations are clustered to produce three structural clusters using the pairwise RMSD between frames as a metric comparing the atoms named CA with a critical distance of 12 Å. Six representative structures from the most populated clusters at 290 K and 350 K are shown in Figure 3.S1. The NAMD energy function were used on these six structures to identify the lowest energy structure, which is shown on Figure 3.S1(b) for 290 K and Figure 3.S1(f) for 350 K. Then, the two chosen single polypeptide structures were placed side-by-side with a 32.3~32.4 Å distance between the centers of mass of the polypeptides and a 7 Å distance between the two closest surfaces. The structures were then solvated in explicit solvent with the closest distance between any solute atom and the edge of the periodic box to be 8 Å and simulated at 290 K and 350 K. The simulation box contained 11,453 water molecules with the size of 5.6 nm × 7.9 nm × 8.3 nm at 290 K and 8,291 water molecules with size of 5.0 nm × 7.0 nm × 8.2 nm at 350 K. Equilibration protocols and MD simulations were the same as described in the section above. The production simulations were performed for 20 ns with a 2 fs time step.
3.3 Results

3.3.1 Conformation of a single (VPGVG)$_{18}$ polypeptide

Our simulations show that in aqueous solution, a single ELP molecule adopts a collapsed state over a temperature range between 290 K and 350 K (Figure 3.1 b-c, Figure 3.2). Previous simulations of the LCST behavior of ELP$^{25}$ and PNIPAM$^{33}$ showed that the value of the radius of gyration ($R_g$) dramatically drops around a critical temperature; we expect that this is most likely an artifact of the short simulation time, particularly at low temperature. In contrast, our simulations of a single (VPGVG)$_{18}$ molecule indicate a gradual decrease in the average value of $R_g$ as the temperature increases (Figure 3.2a). The temporal evolution of $R_g$ at various temperatures is displayed in Figure S3 and indicates that it takes much longer for a polypeptide to assume a stable conformation at lower temperatures than that at high temperatures.

The change in $R_g$ is related to the change in interactions between chemical groups in the polypeptide and the solvent — water— molecules. To examine the contribution of H-bonding at different temperatures, we examine the relationship between the number of polypeptide-water H-bonds, $n_{pw}$, and the number of waters of hydration $N_w$ in Fig 2b. The number of hydration waters is defined as the number of water molecules within a 3.1 Å
distance of any atom on the peptide’s backbone. This distance is taken to be the distance at which the radial distribution function (RDF) between the oxygen of water and an atom on the peptide backbone has its second minimum (Fig. 3.S4). Our simulations reveal that as the temperature increases, both the number of hydrating water molecules and the number of peptide-water hydrogen bonds decreases (Figure 3.2). Moreover, the presence of two distinct clusters one above and one below 330 K (Fig. 3.2b) indicates that there is a strong temperature dependence of peptide-water interactions properties.

The number of intra-peptide hydrogen bonds (Figure 3.2c) with \(n_{p-w-p}\) and without \(n_{pp}\) mediated water demonstrates that as the temperature increases the intra-molecular peptide hydrogen bonding tends to increase and water-mediated hydrogen bonding decreases. Our observation of the temperature-induced increase in intra-molecular peptide bonds agrees well with the experimental observation that peptides at high temperature have more \(\beta\)-turns or \(\beta\)-strands\(^{21, 34}\). The decrease in the number of water-mediated intra-molecular peptide hydrogen bonds also correlates well with the possible expulsion of localized waters at higher temperatures. Overall, the changes in the peptide hydration and the strength of intra-molecular peptide interactions become accentuated in a critical temperature window (i.e., at around 330 K ~ 335 K in our model polypeptide).
3.3.2 Secondary structure analysis

In order to examine the structural propensities of \((VPGVG)_{18}\) as a function of temperature (Figure 3.3) we utilized the DSSP method\(^{35}\), which is a widely used structure recognition algorithm mainly based on H-bonding patterns. In Figure 3, β-turns are defined as having hydrogen bonds between residues \(i\) and \(i+n\), where \(n = 3, 4, 5\) (Fig. 3.3a) and β-strands are defined to include β-bridges and parallel and anti-parallel β-sheets (Fig. 3.3b). We found that as the temperature increases, more β-strands form, yet the number of turns in a single polypeptide shows a non-monotonic change. At low temperatures, it is possible that turns are likely to be stabilized by hydrogen bonding due to the low thermal energy. Increasing the temperature and hence the thermal energy may decrease the stability of intra-chain hydrogen bonding which may lead to a decrease in the number of turn structures at higher temperatures. We observed that the number of turns increases again at even higher temperatures when H-bonding with water decreases and intramolecular hydrogen bonding increases, which is possibly due to the loss of hydration at high temperatures.

The occurrence frequency of secondary structural motifs for each residue is presented in Fig. S8. We observe that Pro and Gly residues involved in turn formation are highly dynamic and undergo rapid interconversion.\(^{36}\) A small number of helical structures form at
the polypeptide termini at all temperatures (Fig. 3.S8). Overall, our observations of
temperature-dependent secondary structure propensities are in general agreement with the
previously discussed experimental observations.\textsuperscript{21-22}

In order to compare our observations with available NMR data we calculated the
torsion angles for each type of amino acid (Val1, Pro2, Gly3, Val4, and Gly5) in (VPGVG)\textsuperscript{18}
at low (290 K) and high (350 K) temperatures (Figure 3.4). We compare our results to the
SS-NMR data for the VPGVG segment in (VPGVG)\textsuperscript{3,22} and [(VPGVG)\textsubscript{4}(VVPKG)\textsubscript{3}]\textsuperscript{16e}
obtained by Hong et.al, and the SS-NMR data for the middle segment in (VPGVG)\textsuperscript{6,19}
obtained by Ohgo et.al. These studies indicate a significant structural heterogeneity and
neither study supported the torsion angles of the β-spiral.\textsuperscript{18} Although there is a reasonable
agreement between these studies on the overall structural propensities of the VPGVG subunit, the observed differences are primarily in the torsion angles for the valine residues in
the VPGVG repeat unit.

Our study indicates that the distribution of dihedral angles is a function of
temperature and that this temperature sensitivity varies from residue to residue in the
pentapeptide motif. The Pro2 and Gly3 are the central two residues, i+1 and i+2, of a β-turn
structure. In Figure 4, the regions represented by labels b1_1 and b2_1 include the torsion
angle pairs for the i+1 residue of a type-II β-turn ($\phi=-60^\circ$, $\psi=120^\circ$). Our results indicate that in general proline in (VPGVG)$_{18}$ tends to adopt the torsion angles for the type II β-turn, a polyproline helix ($\phi=-75^\circ$, $\psi=150^\circ$) or a collagen helix ($\phi\sim=-75^\circ$, $\psi\sim=160^\circ$)). At high temperature (350 K), proline can also adopt torsion angles around the point b2_2 ($-65.4^\circ$, $-23.02^\circ$), which is consistent with the torsion angles for the i+1 residue of a type I β-turn. Ohgo et.al also observed Pro2 angles near ($-60^\circ$, $-30^\circ$) by NMR on the middle segment of (VPGVG)$_6$. 19

As expected, Gly3 can adopt many different conformations. However, within (VPGVG)$_{18}$, it preferentially adopts conformations in the regions marked as c1_1 ($\phi=83^\circ$, $\psi=2^\circ$) at 290 K and c2_1 ($\phi=83^\circ$, $\psi=-10^\circ$) at 350 K. Since the torsion angles of i+2 residues in type-II β-turns are $\phi=80^\circ$ and $\psi=0^\circ$, the Pro2-Gly3 pair favors the type II β-turn structures at low and high temperatures. At high temperature Gly3 also adopts torsion angles labelled as c2_2 ($\phi=-70^\circ$, $\psi=-9^\circ$), which are consistent with i+2 residue in type-I β-turn ($\phi=-80^\circ$, $\psi=0^\circ$) structures. The data for Pro2 and Gly3 indicate that (VPGVG)$_{18}$ adopts a number of type I β-turn structures at high temperatures in addition to the expected type II β-turn structures.
The highest intensity regions for valines indicate that both Val1 and Val4 tend to adopt β-strand conformation, which corresponds to torsion angles of about $\phi = -130^\circ$ and $\psi = 150^\circ$. Val1 conformations are more restricted than Val4, due to the steric hindrance from the pyrrolidine ring of proline. In a solid state-NMR study by Yao and Hong$^{22}$, the torsion angles of the Val1 residue were reported to be $(-96^\circ, 145^\circ)$ or $(-144^\circ, 145^\circ)$, which agrees well with our results. For Val1 the a2_2 ($-70^\circ, 140^\circ$) region has higher intensity at 350 K than that at 290 K which indicate that Val1 has propensity to form more β-turn structures at high temperature. In contrast with Val1, Val4 torsion angles are distributed between four regions where $\phi$ can be $-130^\circ$ or $-90^\circ$ and $\psi$ can be $150^\circ$ or $-5^\circ$. The experimental studies by Yao and Hong$^{22}$ and Ohgo et.al.$^{19}$, which disagree with each other, in combination show four different angle propensities for Val4, which are $(-148,145), (-92, 145), (-110,130)$ and $(-75,-15)^{19}$. These four high intensity regions roughly agree with our observation. Moreover, we show that at high temperatures, Val4 preferentially adopts β-strand structures which is indicated by the higher intensity of $\phi$ angles around $-130^\circ$ (d2_1) in combination with the reduced intensity of d2_3 peak at 350 K compared to 290 K.

The Ramachandran plots for Gly5 and Gly3 are very similar at low temperatures and indicate their propensity to form β-turns. Both glycine residues at 290 K have similar torsion
angles with the highest occupied region at around $(80^\circ, 0^\circ)$, which corresponds to the torsion angle of $i+2$ residues in type II β-turns. However, their angle distributions show different structural preferences at high temperatures. For Gly5, the most populated region at low temperature is around $e_{1\_2}$ of $(81^\circ, -3^\circ)$ and at high temperature is around $e_{2\_1}$ of $(-88^\circ, 148^\circ)$. Thus, at high temperature Gly5 tend to preferentially adopt an extended conformation ($\phi=150^\circ$). Overall, the observed increase in β-strands at high temperatures (Figure 3b) may be attributed to the conformational changes of Val1, Val4 and Gly5. Based on the dihedral angles obtained from the highest intensity points in Ramachandran plots (Figure 4), the representative structures of G(VPGVG)V segment at 290 K and 350 K are built and shown in Figure S9.

3.3.3 Structure of hydration water

The structural properties of water close to the surface of the polypeptide can be quantified by the RDF between the polypeptide and water (Figure 3.5a, Figure 3.S5, and Table S2). We observed that the locations of the first ($r_{\text{min}}=2.23 \text{ Å}$) and second ($r_{\text{sec}}=3.05 \text{ Å}$) local RDF minima are essentially independent of temperature and are primarily due to water-backbone interactions (Figure 3.S5). There is a slight gap in the RDF peaks intensity between the temperatures of 330 K and 335 K, which corresponds to the same temperature window
where change in the ELPs conformational properties has occurred. An increase in temperature leads to a gradual loss of local hydration layers, as indicated by the decrease in the heights of the hydration peaks at higher temperature (Figure 3.5a, Figure 3.S5).

The existence of a hydration shell around the polypeptide is considered to be essential for the LCST behavior of (VPGVG)$_n$ ELPs. The state of the hydrogen-bonded water network in the hydration shell of the short peptide GVG(VPGVG)$_3$ at various temperatures was studied by percolation analysis, where the thermal breaking of the spanning network of the hydration water upon heating was observed. It was proposed that the thermal breaking of the H-bonded network can cause rapid changes of the thermodynamic properties of hydration water which make the surface of the peptide effectively more hydrophobic.

In our simulations, the 400 water molecules that are closest to the polypeptide—which includes the first and second hydration layers—are located within 3.23 Å from the peptide atoms at 290 K and expand to a distance of 3.5 Å at 350 K. Two water molecules were considered as hydrogen bonded when $R_{OO} \leq 3.5\text{Å}$ and $\varphi_{O1-O2-H2} \leq 30^\circ$. We chose this criterion from the simulations of pure water at various temperatures, where each water molecule experiences an average of about 2.24 H-bonds at 300 K and 2.05 H-bonds at 350 K. The size of the largest H-bonded water network in the closest water shell was used to
estimate the connectivity between the water molecules around the ELP. A total of $1.6 \times 10^4$ configurations after equilibrium were examined to identify the largest network of water molecules at each temperature. Probability distribution of the largest water network around the polypeptide at $T=290$ K, 310 K and 350 K are displayed in Figure 3.5b. The profile at lower temperature not only has a larger average network size, but also is broader than the one at higher temperature. Thus, at lower temperature, hydration water molecules appear to be more ordered than at higher temperatures, which corresponds to a decrease in both enthalpy and entropy of hydration water molecules.\textsuperscript{37a, 38} In order to determine the effect of the peptide presence on the network formation of water molecules we compared the water network distributions with and without the peptide. Generally, the temperature increase in bulk water will also reduce the size of the water network (Figure 3.5c, Figure 3.S6). However, the presence of the peptide clearly changes the size distribution of these clusters. Specifically, at high temperatures the presence of a peptide induces the formation of many small water clusters, which is indicated by the narrow distribution as compared to the one for bulk water.
3.3.4 Solvent accessible surface area and energy

To learn what role each amino acid plays in the VPGVG repeat, the interaction energies of each amino acid residue in the repeat unit of the ELP with water molecules were calculated (Fig. 3.6a). The interactions between each amino acid and water molecules weakens as the temperature is increased. Interestingly, proline interacts with water more strongly than other amino acids, regardless of the temperature (Fig. 3.6a). Also, Val1 is generally more hydrophobic than Val4 possibly due to the steric restriction by Pro2.

The solvent accessible surface area (SASA) is another important element in the analysis of protein folding and protein-protein interactions. The concept of accessible surface area also provides a convenient way to define the peptide surface and interior. The SASA is not just a geometric measure, but has physical significance as a gain in hydrophobic interaction free energy is accompanied by a reduction in the SASA. The SASAs of the polypeptide backbone, side chain and various chemical groups are listed in Table S3. The percentage of buried surface area for (VPGVG)$_{18}$ is between 43% and 51% of the SASA of extended polypeptide over the studied temperature range (Table 3.S3). This fraction of buried SASA is slightly lower than that of a globular 100-residue protein, where ~55% of the polypeptide accessible surface becomes buried upon folding. This can be explained by the
lack of a hydrophobic core in VPGVG-ELPs due to the regular distribution of non-polar side chain of each pentamer.

In order to provide a quantitative relationship between the SASA of side chains and the temperature, the SASA fraction associated with the valine and proline side chains and the SASA fraction for glycines are plotted in Figure 3.6b. We observed that as the temperature increases the SASA fraction of for the Val1 side chain and for Gly5 increases, which also correlates with the changes in their torsional angles. This can be explained by an increase in the local turn structures of Val1 and Gly5 at high temperature, which leads to a reduction in structural heterogeneity of the chain.

Our simulations suggest that the LCST behavior of (VPGVG)$_{18}$ may be associated with a abrupt decrease in SASA (Figure 3.6c) which is correlated with the behavior of Val1 and Pro2 sidechains. Specifically, as the temperature increase more side chains of Val1 residues becomes accessible for water and Pro2 residues are hidden from water. Generally the side-chain mobility increases as the temperature increases, however in our case the observed reduction in ELP radius of gyration reduces the side-chain motion. The observed thermo-responsive changes in SASA of Val1 and Pro2 side chains in the polypeptide are due
to the competition between these two effects. Our simulation also suggests that increasing temperature leads to less heterogeneity of the peptide structure.

### 3.3.5 Polypeptide aggregation

The temperature-triggered coacervation of ELPs has been described as a concentration-dependent phase separation,\(^3\) where hydrophobic intermolecular association between peptides becomes the dominant process. The formation of microscale aggregates by systems containing multiple polypeptide chains occurs on a time scale that is not readily accessible in atomistic MD simulations. Therefore, to examine whether the temperature-dependent changes at a single peptide level will lead to aggregation, we conducted simulations of two temperature-equilibrated polypeptides at low temperature (290 K) and at high temperature (350 K). Initially, the polypeptides, despite being spatially close, lack the necessary molecular contacts that would bias their aggregation into a compact assemble. This simulation was performed to reveal the early stage of polypeptide aggregation. Figure 3.1 show snapshots of the two polypeptides at 290 K and 350 K after 15 ns of simulation time. At high temperature (350 K), the two polypeptides form a strongly-compacted aggregate (Figure 3.1e), with water expelled from the interface between the two molecules. At low temperature (290 K), the polypeptides are in a much more extended conformation and can
intermittently interact with each other due to the periodic boundary conditions (Fig. 3.1d). Thus, upon transition from low to high temperature, the system undergoes a transition from a disordered state to an aggregate with a higher degree of compactness. The distance between the center-of-mass of the two chains is displayed in Figure 3.7a to show the dynamics of aggregation. The interaction energy profiles between the polypeptides for both cases (Figure 3.7b) correlates with the distance between their centers of mass, and shows that the chains interact strongly at high temperatures. During the simulations, no significant changes in the radius of gyration and secondary structure propensities were observed (Figure 3.S10).

The increase in hydrophobicity with temperature of a single peptide may play a key role in the process of aggregation. The correlation between the interaction energies between water and the peptide and between the two polypeptides is plotted on Figure 8. The temperature increase leads to a reduction in water-peptide interactions (i.e. increase in hydrophobicity of a single polypeptide) and consequently to a strong association between the polypeptides. A single polypeptide molecule in water is more stable at low temperature than high temperature, while a polypeptide aggregate is less stable at low temperature than at high temperature.
3.4 Conclusions

Using long atomistic molecular dynamics simulations of \((VPGVG)_{18}\) in aqueous solutions, we investigated the temperature-dependent structure and dynamics of the polypeptide chain and its hydration water, in order to understand the mechanisms underlying the LCST behavior of VPGVG polypeptides. Based on our simulations, we conclude that the transition behavior includes thermal disruption of the water network, loss of hydration of the polypeptide, and an increase in its hydrophobicity, leading to physical association of chains.

Our simulations show that at high temperatures, an ELP molecule tends to form a more ordered secondary structure in the form of \(\beta\)-strands. Detailed analysis of residue-based torsion angles in the VPGVG repeat unit shows that \(\beta\)-turn structures exist at all temperatures. Specifically, the Pro2-Gly3 pair can adopt type II \(\beta\)-turn structures at all temperatures. At high temperatures there is also a number of type I \(\beta\)-turns formed for Pro2-Gly3. In contrast, Val1 adopts \(\beta\)-strand conformations at all temperatures and more turn conformations at high temperatures. However, the formation of \(\beta\)-strand structures becomes more prominent for Val4 and Gly5 at high temperatures. The presence of prolines and glycines generally limits the formation of extended secondary structures due to the steric constraints imposed by prolines and the high entropic penalty of glycine confinement.\(^{34, 42}\)
However, this limitation can be compromised when the temperature is raised\textsuperscript{1b}, which allow for more β-strand and distorted β-strand structures to be adopted at high temperatures.

At high temperatures we observed thermal disruption of the hydrogen-bonded water network around the polypeptide and an abrupt decrease in the polypeptide’s SASA, which are accompanied by increases in secondary structure formation and exposure of hydrophobic side chains as well as displacement of local hydrating water molecules into the bulk. From the traditional peptide backbone centric view, where the peptide backbone is considered responsible for the resultant conformation, the role of hydrophobicity is non-specific. However, from the peptide side-chain centric view, hydrophobicity is the dominant force for protein folding.\textsuperscript{43} Moreover, proline in elastomeric proteins is considered to be a “gatekeeper” residue which maintains the disordered structure and prevents collapse into a hydrophobic core.\textsuperscript{1b, 34, 44} Interestingly we observed a distinct change in the exposure of Val1 and Pro2 hydrophobic side chains at high temperatures. Thus, as the temperature increases, ELPs undergo conformational changes of the backbone along with the hydrophobicity changes of the side chains, exposing hydrophobic valine-side chain to the solvent and hiding the “gatekeeper” proline residue.
Simulations of the initial stages of the aggregation process between two polypeptides demonstrated that the observed gradual changes in the properties of a single polypeptide with temperature are responsible for the experimentally observed aggregation of ELPs at or above LCST. Based on analysis of the interaction energy, we suggest that the competition between peptide-peptide and peptide-water interactions may determine the LCST of the system. Overall, we propose that the LCST in VPGVG polypeptides is a collective phenomenon that originates from gradual changes in the structure of single polypeptide chains and the abrupt change in properties of hydration water around the peptide as temperature increases. Because we only observed gradual structural changes with temperature at the single peptide level, this new molecular level understanding of the structure, dynamics, and thermodynamics of ELP (VPGVG)$_{18}$ as a function of temperature reinforces the concept of cooperative nature of the pronounced phase transition exhibited by VPGVG polypeptides.

Supporting Information Available: Additional information about the simulation systems, convergence of simulations and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*yara_yingling@ncsu.edu

Funding Sources

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107).

ACKNOWLEDGMENT

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107).

The computer support was provided by the High Performance Computing Center at North Carolina State University.
3.5 References


Figure 3.1 (a) Side and front view of the β-spiral structure for (VPGVG)$_{18}$. Final simulation snapshots of (b,c) a single (VPGVG)$_{18}$ structure and (d,e) the interaction between two ELPs at (b,d) 290 K and (c,e) 350 K. The backbone is represented as secondary structure ribbons where turn motif is colored in green, β-sheet is yellow and coil is silver. All heavy atoms are depicted as sticks. In panel a-c, the side chains are colored by amino acid (Gly in purple, Pro in yellow and Val in blue). In panel d-e, the side chains are colored by molecule.
Figure 3.2 (a) Temperature dependence of the radius of gyration ($R_g$) of (VPGVG)$_{18}$. (b) Correlation between $N_w$, the number of water molecules in the second watershell of backbone, and $n_{pw}$, the number of peptide-water hydrogen bonds. The temperature is labeled at the data points. (c) Number of the intramolecular intra-peptide hydrogen bonds (black) and intramolecular water-mediated peptide hydrogen bonds (red). The error bars in these plots represent the standard deviation.
Figure 3.3 Secondary structure formation as a function of temperature for (a) β-turn and (b) β-strand structures.
Figure 3.4 Ramachandran plot ($\phi$, $\psi$ distributions) for each residue in a pentamer, (a1,2) Val1, (b1,2) Pro2, (c1,2) Gly3, (d1,2) Val4, (e1,2) Gly5, at two temperatures, 290 K (left column) and 350 K (right column), colored by intensities. The most populated regions on the Ramachandran plots are labelled.
Figure 3.5 (a) Radial distribution function (RDF) of oxygen atoms in water molecules around polypeptide atoms as a function of temperature. (b) Probability distribution of the size $N_{\text{max}}$ of the largest water network around polypeptide. (c) Probability distribution of the size $N_{\text{max}}^W$ of the largest water network in pure water systems.
Figure 3.6 Temperature dependence of properties of a single ELP chain: (a) Interaction energy of each individual residue in an ELP pentamer with water. (b) Fraction of SASA of Val1_side chain (black), Pro2_side chain (blue), Val4_side chain (red), Gly3 (green) and Gly5 (pink). (c) SASA of peptides (black; square) and SASA fraction of Val1_side chain (blue; triangle) and Pro2_side chain (red; circle).
Figure 3.7 Interaction between two ELPs. (a) Distance between center-of-mass of two polypeptides and (b) peptide-peptide interaction energy at 290 K and 350 K.
Figure 3.8 Peptide-water interaction energy in single-peptide system (black; square). Peptide-peptide interaction energy in double-peptide system (blue; triangle). The data are averaged over the last 40 ns for each MD run in a single chain system and over the last 5 ns run for double chain system.
The physical origin of the lower critical solution temperature (LCST) behavior of a variety of fluids, including elastin-like polypeptides (ELPs), has been studied for the past few decades. As is the case for polymer solutions, LCST behavior of ELPs is invariably reported for large systems of molecules and is considered evidence for collective behavior. In contrast, we find evidence for properties changes associated with LCST behavior in a single molecule by performing long atomic-level molecular dynamics simulation on the ELP sequences (Val-Pro-Gly-Val-Gly)n for four different length peptides over a wide range of temperatures. We observe a sharp transition in the number of hydrogen bonds between peptide and water and in the number of water molecules within the first hydration shell as temperature rises; this is used to locate the transition temperature. The dependence of the transition temperatures of ELPs on their lengths agrees well with experiments in that both have the same power law exponents. Our simulations reveal that the tendency for pentamers (VPGVG) in ELPs of all lengths to lose H-bonds with water or to gain H-bonds with themselves as temperature rises is independent of the length of the chain in which they are
embedded. Thus, the transition temperature of ELPs in pure water is determined by two factors: the hydrogen bonding tendency of the pentamers and the number of pentamers per ELP. Moreover, the hydrogen bonding tendency of pentamers depends only on their sequences, not on the ELP chain length.

4.1 Introduction

Elastin-like polypeptides (ELPs) are artificial polypeptides having sequences derived from the hydrophobic domains of tropoelastin, elastin’s precursor. ELPs consist of a pentapeptide repeat, Val-Pro-Gly-Xaa-Gly, where Xaa, the guest residue, can be any natural amino acid except proline. ELPs exhibit lower critical solution temperature (LCST) transition behavior, which means they are soluble in aqueous solutions when the system temperature is below the LCST and insoluble when the temperature is above the LCST, resulting in a polymer-rich phase and a polymer-poor phase. The transition temperature \( T_C \) of ELPs depends not only on the solution composition (the lowest value being the LCST) but also on their properties such as chain length, and the identity and composition of the guest residues, which can be precisely controlled. The transition temperature of ELPs can be modified by physiochemical means such as changes in pressure, salt concentration, and pH, etc. All these stimuli-responsive features as well as their low toxicity compared to other polymers make ELPs potential candidates in biological applications, such as drug delivery carriers, molecular switches, tissue engineering scaffolds and purification tools for recombinant protein expression.
The physical origin of LCST behavior and its dependence on chain length have been the subject of numerous studies over the last few decades. For polymers such as poly-(ethylene oxide) (PEO)\textsuperscript{30,31} and poly(N-isopropylacrylamide) (PNIPAM)\textsuperscript{32,33}, the transition temperature $T_\text{t}$ decreases as chain length increases; LCST behavior is not observed when chain length is too short. Theory-based studies of the length dependence of LCSTs for model polymers also indicate that polymers with chain lengths of seven segments or more exhibit LCST behavior but those with less do not\textsuperscript{34}. The consensus regarding the explanation for LCST behavior has been that LCST behavior for polymer solutions is a consequence of compressibility differences between solute and solvent which get accentuated as the solute lengthens, but that LCST behavior for smaller molecules is due to the existence of directional forces\textsuperscript{35,36}. ELPs fall in between these two cases, they are relatively long and they do experience directional interactions.

Systematic approaches to understanding the physical origins of LCST behavior in ELPs and the length dependence of the transition temperature are still lacking. Elucidation of these matters would enable us to gain better control of the thermal properties of ELPs, meaning we could create ELPs with desired transition temperatures, and hence desired transport phenomena and biodegradation capability, etc. Although the LCST behavior of ELP has been mainly attributed to hydrophobic effects\textsuperscript{37}, Cho et al. found that hydrogen bonding also plays an important role in stabilizing the collapsed state of ELP in heavy water D$_2$O\textsuperscript{38}. Ma et al.\textsuperscript{39} found a change of slope for (VPGVG)$_3$ at around 30 °C in the NMR relaxation time of the ELP heating curve which was proposed to correlate to the inverse
Zeeshan et al.\textsuperscript{40} compared the conformational difference between cyclic (VPGVG)$_3$ and linear GVG(VPGVG)$_2$ ELP sequences and reported a decrease in the global minimum of the CD spectra with increasing temperature, which they regarded as evidence of ELPs inverse temperature transition. Meyer et al\textsuperscript{5} conducted experiments on [V$_i$A$_j$G$_k$]$_n$, sequences consisting of a total number of n pentamers, Val-Pro-Gly-Xaa-Gly, where Xaa = V for i pentamers, A for j pentamers and G for k pentamers in each repeating unit to quantify the effect of chain length on the transition temperature, $T_t$, and reported that $T_t \propto (\text{length})^{-1}$ at constant ELP concentration. Hest et al.\textsuperscript{41} measured the $T_t$ of ELP sequences [V$_i$L$_j$G$_k$]$_n$ with different molecular weights and showed that $T_t$ decreases with increasing chain length, which agrees with Meyer et al.\textsuperscript{5} and is the same trend as for PNIPAM\textsuperscript{32,33}. Based on these trends one would expect that the short ELP sequences such (VPGVG)$_3$ should have a much higher transition temperature than the 30 °C reported by Ma et al.\textsuperscript{39}. Nicolini et al.\textsuperscript{9} reported that a short sequence, GVG(VPGVG), had an inverse transition temperature around 36 °C in D$_2$O using FTIR measurements; however, the use of D$_2$O as a solvent could affect the LCST behavior as reported by Cho et al.\textsuperscript{38} In contrast Nuhn and Klok\textsuperscript{42} investigated the effects of chain length on ELP sequences based on the GVGVP motif and observed that the LCST value decreases as the chain length increases which is in agreement with Meyer et al.\textsuperscript{5} and Hest et al.\textsuperscript{41} observations. However, they did not observe an LCST of (GVGVP)$_n$ sequence when the chain length was less than 4. Overall, while the longer ELP chains clearly exhibit length-dependent transition temperatures, the existence of LCST for ELPs shorter than n = 4 remains a controversy.
Molecular dynamics (MD) simulations have been performed to elucidate the temperature-induced conformational changes of ELP sequences of different chain lengths. Rousseau et al.\textsuperscript{43} proposed that the LCST of GVG(VPGVG) is between 313 K and 333 K, which was based on observations that while the extended structural conformations predominate at all temperatures some compact structures are observed at high temperatures.

In 20ns simulations of (VPGVG)\textsubscript{3} and (LGGVG)\textsubscript{3}, and 4ns simulations of (VPGVG)\textsubscript{7} and (LGGVG)\textsubscript{7}, Huang et al.\textsuperscript{44} observed that the trends in the number of peptide-peptide hydrogen bonds and the number of peptide-water hydrogen bonds changed as the temperature increased which they attributed to an inverse temperature transition. Ma et al.\textsuperscript{39} performed 4 ns MD simulations on (VPGVG)\textsubscript{3} and reported that the number of water molecules around putative deuterium-labeled sites in their experimental study changes as temperature increases. They attributed this to a folding transition at a temperature of around 30 °C, meaning that the structure becomes more ordered when the temperature is higher than the transition temperature and more flexible at temperatures lower than the transition temperature. However, very long simulations (up to 350 ns) by Krukau et al.\textsuperscript{45} conducted on GVG(VPGVG)\textsubscript{3} challenged the observation of an inverse temperature transition for such short chains. They showed that this ELP is highly flexible at high temperatures and becomes more rigid at low temperatures, which is the opposite of what Huang et al. concluded. Thus, the existence of a transition temperature for short ELPs is still uncertain.

MD simulations of longer chains, which are well known to have LCSTs, are able to shine more light on the origin of inverse transition temperature in ELPs. Simulations of a
much longer chain \((VPGVG)_{18}\) by Li et al.\(^{46}\) did not find evidence for a particular transition temperature perhaps because their simulation at 10 ns, was not very long; being limited by the computational power at that time. In a previous paper, Li et al.\(^{47}\) reported MD simulations results for \((VPGVG)_{18}\), and found \(T_t\) for a single peptide to be between 330 K and 335 K based on analysis of hydration water properties. The results from that paper demonstrated that the LCST phenomenon originates from an abrupt change in the hydration properties of ELP peptides as temperature rises. These results inspired us to perform explicit solvent atomistic molecular dynamics simulations of \((VPGVG)\) ELPs of different lengths to see if these hydration property changes were general and how they depend on chain lengths.

In this study, we simulate ELPs of four different lengths: \((GVG)(VPGVG)_{3}\), \((VPGVG)_{10}\), \((VPGVG)_{18}\) and \((VPGVG)_{30}\) using atomistic, explicit-solvent MD simulations over a wide range of temperatures from 285 K to 490 K. Our motivation is not only to pinpoint the length dependence of the LCST behavior of ELPs, but also to use this study as a vehicle to better understand the origin of their LCST behavior and to get a full picture of what drives LCST phenomenon on a molecular level. Here, we report a comprehensive investigation of temperature and length dependent properties of ELP which shows that a single ELP molecule also exhibits behavior related to LCSTs. Two factors determine the transition temperature behavior of ELPs, the hydrogen bonding tendency of the individual pentamers and the number of pentamers in the ELP. Moreover, the hydrogen bonding tendency of each pentamer depends only on its sequence, not on the length of the chain in which it is embedded.
4.2 Materials and Methods

4.2.1 ELP simulations with single chain.

We performed explicit-solvent, atomistic simulations on (GVG)(VPGVG)$_3$, (VPGVG)$_{10}$, (VPGVG)$_{18}$, (VPGVG)$_{30}$ using Amber 11 with ff99SB force $^{48}$. Urry’s β spiral model $^{49}$ was used as the initial configuration for peptides of all lengths. The peptide was first solvated in the box with explicit TIP3P water molecules $^{50}$. Based on our experience in a previous study $^{47}$ on (VPGVG)$_{18}$, we chose to conduct our simulations over a wide range of temperatures: 290 K-490 K for (GVG)(VPGVG)$_3$, 300 K-375 K for (VPGVG)$_{10}$, 290 K-350 K for (VPGVG)$_{18}$ and 285 K-350 K for (VPGVG)$_{30}$. The simulation procedures described below were applied to peptides of all different lengths.

Each simulation contained the following stages. First, a 10,000-step solvent minimization was performed using the steepest descent method with the peptide restrained via a force of 200 kcal/mol. Then the system was gradually heated over 10 ps to a target temperature with the peptide restrained via a force of 200 kcal/mol, after which a short 40ps NPT ensemble MD run was conducted, again with the peptide constrained via a force of 200 kcal/mol. This was followed by another 10,000-step energy minimization process with a 25 kcal/mol restraint. Subsequently another 20 ps NPT ensemble MD run was performed with a force of 25 kcal/mol. Then another 10,000-step unconstrained minimization was performed followed by a reheating over 40 ps of the system at a constant volume to the target temperature. Finally, an NPT ensemble MD run was performed to ensure uniform solvent density. The Berendsen thermostat $^{51}$ was used to maintain the simulation temperature.
Partial mesh Ewald (PME) \(^{52}\) summation was used to calculate the long-ranged electrostatic interactions and a 9 Å cut-off radius and a 1E-5 tolerance for the Ewald convergence were used to calculate the non-bonded interaction. The SHAKE \(^{53}\) algorithm was used to constrain bonds involving hydrogen atoms. The run times for the simulations were 50 ns for (GVG)(VPGVG)\(_3\), 60 ns for (VPGVG)\(_{10}\), 70 ns for (VPGVG)\(_{18}\) and 160 ns for (VPGVG)\(_{30}\). The convergence was confirmed by the data presented in the Supporting Information (Figure 4.S3 (a-d)). The last 30-40 ns of the trajectories from each case were considered for statistical analysis. Amber tools, VMD, NAMD and in-house scripts were used for statistical analysis.

4.2.2 Hydrogen bonding analysis.

The hydrogen bonding analysis was performed with a distance cut-off of 3.5 Å and an angle cut-off of 30°. The number of water molecules in the first hydration shell of the peptide backbone, \(N_w\), was determined by counting the number of water molecules within 2.43 Å of the peptide. This distance corresponded to the first minimum in the radial distribution function (RDF) between the oxygen atoms on the water molecules and the backbone atoms on the peptide (see Figure 4.S4).

4.2.3 Secondary structure analysis and interaction energy calculation.

Secondary structures were determined using the DSSP algorithm \(^{54}\), which has been widely used as a standard method for assigning secondary structures to amino acids. In our study we only considered β turn and β strand structures including parallel, anti-parallel and β bridge structures. The β turn structure was defined to be a turn-like structure formed by residue i and i+n where (n = 3, 4, 5). Interaction energies between peptide and water, and
between peptide and peptide, were calculated with NAMD. These include only Van der Waals (VdW) and electrostatic (ELE) energies.

4.3 Results

LCST transition is exhibited by a single chain based on analysis of hydration properties. We analyzed the number of hydrogen bonds (H-bonds) between peptide and water, $N_{pw}$, versus the number of water molecules within the first hydration shell of the peptide backbone, $N_w$ (Figure 4.1). The criteria for calculating these two parameters are described in the materials and methods section. The length of each ELP sequence, $L$, is defined to be the number of pentamers in each sequence, e.g. $(VPGVG)_30$ has length, $L = 30$, and $(GVG)(VPGVG)_3$ has length, $L = 3.5$.

The overall trend of the data in $N_{pw}$-$N_w$ space is that as temperature rises, both $N_{pw}$ and $N_w$ decrease, indicating less peptide-water interactions. We also observed that for peptides with length $L=30$, the data points in $N_{pw}$-$N_w$ space are grouped into two distinct clusters as seen in Figure 4.1A, indicating a clear temperature-induced transition in the hydration properties for a single peptide. This suggests that a single ELP $(VPGVG)_n$ molecule exhibits an LCST-type transition temperature ($T_t$) at the points in $N_{pw}$-$N_w$ between two temperature-based clusters. Thus the transition temperature for $(VPGVG)_{30}$ is $307.5 \pm 2.5$ K, $T_t$ for $(VPGVG)_{18}$ is $332.5 \pm 2.5$ K, and $T_t$ for $(VPGVG)_{10}$ is $345 \pm 5$ K. As we can see from Figure 1, the separation between the two clusters diminishes as chain length decreases; there is no transition when the chain length is too short ($L=3.5$). This decrease in $T_t$ with increasing
chain length is the same as that observed in experiments conducted by Meyer et al.\textsuperscript{5}, who found that $T_t$ for ELP sequence (VPGVG)$_n$ decreases as chain length increases. Meanwhile, our finding that very short ELPs don’t exhibit LCST transitions is supported by experimental observations by Nuhn et al.\textsuperscript{42} who found that ELP sequence (GVGVP) does not have an LCST transition temperature when the chain length is shorter than 4 ($L<4$).

Transition temperatures from simulations and experiments have the same power law dependence on length. The transition temperature versus chain length data observed in our simulations can be compared with experimental data by fitting both sets to power law expressions as shown in Figure 4.2. Experimental data on $T_t$ vs. $L$ was obtained from Meyer et al.\textsuperscript{5} for a series of ELP (VPGVG)$_n$ sequences with an ELP concentration of 25 μM in phosphate buffered saline solution. Fitting their experimental and our simulation $T_t$ values with a power law function $T_t = c \times (\text{length})^\nu$ yields very similar values for the exponents for the two curves: $\nu = -0.65$ for simulation and $\nu = -0.63$ for experiment. There is a shift in the temperatures of the two fitted curves, which could be because of several reasons: the simulations are performed in pure water whereas the experiments were conducted in PBS solutions with different types of salt; there is a difference in ELP concentration between experiment and simulations; the predicted temperatures can be affected by the heating rate or the properties of the water model used in simulations.

Thus, we see that even though the simulations were performed on a single peptide molecule and experiments were performed on multi-peptide systems, there is quantitative agreement on the length dependence of $T_t$ between experiments and simulations. This lends
support to our hypothesis that LCST behavior is reflected in the behavior of a single molecule, meaning that LCST behavior is not only a collective behavior of many molecules, but also an intrinsic property within a single peptide.

The hydrogen bonding tendency of each pentamer depends only on the sequence, not on the ELP chain length for (VPGVG) sequences. Figure 4.3A shows the number of H-bonds between peptide and water, \( N_{pw} \), versus temperature. We can see that \( N_{pw} \) decreases as temperature increases for ELPs of all lengths, indicating that the attraction between peptide and water weakens as temperature increases. The slopes of the linear fits to these curves for each length become more negative as the chain length increases. Thus, the number of H-bonds between chain and water for the longer chains decreases faster as temperature rises than for the shorter chains. One explanation for this is that the longer the chain is, the larger its solvent accessible surface area is, and, hence, the more water molecules it has in its hydration shell. Therefore, compared to the shorter chains, the longer chains have more water molecules to lose from their hydration shells as temperature increases, meaning less water will be available to hydrogen bond with the peptide.

To learn if and how the hydrogen bonding capability of a VPGVG pentamer depends on the length of the chain, we divided the slope of the temperature dependent number of hydrogen bonds between peptide and water (\( N_{pw} \)) by the number of pentamers, which is \( (dN_{pw}(L)/dT)/L \). In the inset of Figure 3(A) we plotted \( (dN_{pw}(L)/dT)/L \) versus chain length \( L \) and show that

\[
\frac{dN_{pw}(L)}{dT} \frac{1}{L} = \frac{dn_{pw}(L)}{dT} = \text{constant} \tag{4.1}
\]
where npw(L) is defined to be Npw(L)/L. Equation 1 shows that each VPGVG pentamer loses the same number of H-bonds with water as temperature increases, and that this is independent of the ELP chain length. It could be that the temperature-dependent H-bonding pattern of ELP with water depends only on the sequence of its basic building block and not on the overall length of the chain. Given the connection between Npw and the transition temperature, the transition temperatures of ELPs of each length could be associated with a collective behavior of all the pentamers within a single chain.

Figure 4.3B shows the number of H-bonds between peptide and peptide, Npp, versus temperature. We can see that Npp increases as temperature increases for L=18 and 30, signifying a stronger intra-molecular attraction for those lengths when temperature rises. However, when L≤10, the peptide-peptide hydrogen bonding seems to be independent of chain length. This is because a short chain is much more flexible than longer ones, which makes it harder to retain peptide-peptide hydrogen bonds. Reference to the inset graph in Figure 4.3(B) and using an analysis similar to that used above, we conclude that

\[
\frac{dN_{pp}(L)}{dT} \frac{1}{L} = \frac{dn_{pp}(L)}{dT} = \text{constant}
\]

where \( n_{pp} \) is the number of intra-molecular hydrogen bonds per pentamer, \( n_{pp}(L)=N_{pp}(L)/L \). This indicates that each pentamer’s tendency to form peptide-peptide hydrogen bonds as temperature increases is approximately the same no matter what the chain length is.
Figure 4.4A shows $\Delta n_{pw}$, the difference between $n_{pw}^{\text{above}}$ (the average of $n_{pw}$ over all temperatures above $T_t$), and $n_{pw}^{\text{below}}$ (the average of $n_{pw}$ over all temperatures below $T_t$) per pentamer for all chain lengths. The negative value for $\Delta n_{pw}$, $-0.93$ (in triangles) indicates that a pentamer will lose approximately 0.93 peptide-water H-bonds when temperature rises from below to above $T_t$. Similarly the positive value for $\Delta n_{pp}$, $0.13$ (in squares) indicates that a pentamer will gain approximately 0.13 peptide-peptide H-bonds when temperature rises from below to above $T_t$. Our finding that $|\Delta n_{pw}|>|\Delta n_{pp}|$ might be because the chain becomes more compact as temperature rises from below $T_t$ to above $T_t$, thus the hydrophobic side-chains have more contacts with each other within the peptide. This makes peptide-peptide hydrogen bonds harder to form; therefore the increase in $n_{pp}$ is not as big as the decrease in $n_{pw}$.

In Figure 4.4B, we present a schematic representation of the configurations adopted by ELP (VPGVG) at temperatures below and above the transition temperature. When $T<T_t$, the chain is more extended and relaxed and has more peptide-water H-bonds (green) and less peptide-peptide H-bonds (magenta) than when $T>T_t$.

The ELP’s other general properties exhibit transition behavior. Table 4.1 presents the values of a number of properties per pentamer calculated during the simulations for peptides of lengths $L = 10, 18$ and $30$. We do not include (GVG)(VPGVG)$_3$ since we didn’t observe a transition for that length in our $N_{pw}-N_w$ analysis. The table has three sections from top to bottom: average values of properties over all temperatures below $T_t$, average values of properties over all temperatures above $T_t$, and their differences. The last line in the third section contains the average of the differences over all lengths.
The solvent accessible surface area (SASA) per pentamer (second column) decreases as chain length becomes larger both below $T_t$ and above $T_t$. This implies that as chain length becomes longer, each pentamer will have less opportunity to contact water molecules. This is consistent with the values for the interaction energy between hydrophobic side chains per pentamer (third column), $E_{hp/penta}$. As chain length increases, the interaction energy between hydrophobic side chains per pentamer increases with temperature both below $T_t$ and above $T_t$, indicating that the pentamer becomes more hydrophobic. As for the difference between the SASA above and below $T_t$, there is a small increase in SASA as temperature rises from below to above $T_t$ for peptides with $L=10$ ($\Delta=11.63$) and a large decrease in SASA for peptides with $L=18$ ($\Delta=-30.80$) and $30$ ($\Delta=-23.82$). The small increase in SASA at $L=10$ might be because the small peptide’s structure fluctuates more at high temperatures than that of longer peptides. In other words it does not stay as compact as longer peptides do when $T>T_t$, thus the SASA when $T>T_t$ appears to be a little bit higher than when $T<T_t$.

We examined how the numbers of $\beta$ turns and $\beta$ strands change with temperature below $T_t$ and above $T_t$. The percentages of $\beta$ turns per pentamer for chains of different lengths below $T_t$ and above $T_t$ (fourth column of the first two sections) decrease as chain length increases, indicating that longer chains adopt less $\beta$ turn structure than shorter ones. The percentages of $\beta$ strands per pentamer (fifth column) also decrease as chain length increases, indicating that the chain is getting less ordered as it become longer. However, the percentage of $\beta$ strands per pentamer increases as temperature goes from below $T_t$ to above $T_t$, i.e. $\Delta \beta_{strands/penta} > 0$ implying that the chain is more ordered above $T_t$ than below $T_t$.  

108
Meanwhile, $\Delta \beta_{\text{strands/penta}}$ also decreases as chain length increases, signifying that it is more difficult for longer chains to form a more ordered structure as temperature goes from below $T_t$ to above $T_t$ than for shorter chains. We examined how the numbers of $\beta$ turns and $\beta$ strands change with temperature below $T_t$ and above $T_t$. The percentages of $\beta$ turns per pentamer for chains of different lengths below $T_t$ and above $T_t$ (fourth column of the first two sections) decrease as chain length increases, indicating that longer chains adopt less $\beta$ turn structure than shorter ones. The percentages of $\beta$ strands per pentamer (fifth column) also decrease as chain length increases, indicating that the chain is getting less ordered as it become longer. However, the percentage of $\beta$ strands per pentamer increases as temperature goes from below $T_t$ to above $T_t$, i.e. $\Delta \beta_{\text{strands/penta}} > 0$ implying that the chain is more ordered above $T_t$ than below $T_t$. Meanwhile, $\Delta \beta_{\text{strands/penta}}$ also decreases as chain length increases, signifying that it is more difficult for longer chains to form a more ordered structure as temperature goes from below $T_t$ to above $T_t$ than for shorter chains. We compare the highest intensity regions on the Ramachandran plot for Val1, Pro2, Gly3, Val4, and Gly5 on ELPs with lengths $L=10$, $18$ and $30$ (Figure 4.S1). We find that the location of high intensity peaks in the Ramachandran plot have a similar pattern at high and low temperatures for all ELPs in this study and are independent of length. At low temperature, Pro2 adopts (−75°, 150°), Gly3 adopts (83°, 2°) and Gly5 adopts (80°, 0°), which is indicative of Pro2, Gly3 and Gly5 adopting type II $\beta$ turn conformation. Val1 adopts (−130°, 150°) and Val4 adopts 4 regions (with $\phi$ around 130° or 90° and $\psi$ around 150° or −5°), which signifies that Val1 and Val4 adopt $\beta$ strand conformations. At high temperature new regions appear for all
five residues. Pro2 adopts (-65°, -23°) and Gly3 adopts (-70°, -9°) which is indicative of Pro2 and Gly3 adopting type I β turn. Val1 adopts (-70°, 140°) and Gly5 adopts (-88°, 148°), which indicates that Val1 adopts more of a β turn and Gly5 adopts more extended structure than at low temperature. At high temperature Val4 has a higher intensity of φ angles of -130° than at low temperature, which shows that Val4 adopts more β strand structure than at low temperature. Our previous study on (VPGVG)$_{18}$ indicated that the simulation-based prediction of the most probable regions on the Ramachandran plot for the different amino acid residues agrees well with the SS-NMR studies on (VPGVG)$_{3}$ and [(VPGVG)$_{4}$(VVPGKG)$_{39}$]$_{56}$ conducted by Hong et al. and the SS-NMR data for (VPGVG)$_{6}$$_{57}$ investigated by Ohgo et al. Thus, the observed most probable regions in Ramachandran plot, which are independent of length in our simulations, are in agreement with experiments.

The values of the interaction energies between pentamer and water ($E_{pw}$/penta) and within peptides ($E_{pp}$/penta) are shown in the sixth column and seventh columns in Table 4.1. The interaction energies (which include the van der Waals (VdW) and electrostatic (ELE) energies) between peptide and water increase (become less negative) as chain length increases both below and above $T_{c}$. As was the case for the SASA, the reason behind this is that as the chain lengthens, more pentamers are buried inside the compact structure, so there is less area accessible for solvent in vicinity of the peptide. This makes for less interaction opportunities between peptide and water, which leads to an increase in the peptide-water interaction energy. The differences between $E_{pw}$/penta as temperature increases from below
to above $T_i$, $\Delta E_{pw}/\text{penta}$, are all positive (fifth column of the third section), meaning that there are less interactions between water and peptide as temperatures rises from below to above $T_i$. The difference between $E_{pp}$ as temperature goes from below to above $T_i$, $\Delta E_{pp}/\text{penta}$, for $L=18$ and $30$ reveals that the peptide interacts with itself more as temperature rises from below to above $T_i$. This is also consistent with our result on intra-peptide hydrogen bonding.

We investigated the size distribution of the largest water clusters among the 400 water molecules closest to the peptide (see Figure 4.S2). The water cluster is defined to be a collection of water molecules that are connected with each other through hydrogen bonding but not with peptide. The distributions for all lengths had larger average water network sizes and were broader at lower temperatures than at high temperatures, meaning more water molecules experience directional forces at low temperatures than at high temperatures. Thus, the hydration water is more ordered at lower temperatures than at higher temperatures.

### 4.4 Discussion

The LCST behavior of ELPs is invariably reported for multiple-chain systems and is considered evidence for collective behavior, but we find evidence for property changes associated with LCST behavior in a single molecule. A clear transition with temperature is observed in ELPs’ hydration properties, more specifically the number of H-bonds between peptide and water and the number of water molecules within the first hydration shell of the peptide backbone. While LCST behavior is certainly a collective phenomenon and the $T_i$ does depend on peptide concentration, our study shows that the seeds for LCST behavior,
that is the competition between water-water, peptide-water and peptide-peptide interactions, are evident at the single molecule level.

We systematically investigated the length dependence of LCST behavior of ELPs based on the pentamer unit VPGVG for four different lengths (L = 3.5, 10, 18 and 30) by analyzing their hydration and structure over long molecular dynamics simulations trajectories. A clear gap is evident in the space spanned by the number of peptide-water hydrogen bonds (N_{pw}) and the number of water molecules within the first hydration shell of the peptide backbone (N_{w}) for all lengths but L = 3.5. This gap, which we identify with the LCST transition, was observed by us previously in simulations of L=18 by Li et al\textsuperscript{47}. We confirm this observation here for two other chain lengths. Thus it appears that a single ELP molecule (VPGVG)\textsubscript{n} exhibits a LCST-type transition behavior. Our finding that chain (GVG)(VPGVG)\textsubscript{3} does not display this behavior is consistent with experimental results on very short ELPs but contrary to simulation observations on GVG(VPGVG)\textsuperscript{43} and (GVG)(VPGVG)\textsubscript{3}\textsuperscript{45}, which might be related to the length of the simulations.

Our simulation results agree well with experimental results regarding the trend in the transition temperature with chain length. When the T\textsubscript{T} versus length curves are fit to a simple power law, the exponents from simulations and experiments are remarkably close. The different prefactors of the two curves can be explained by a difference in the conditions at which the simulations and experiments are conducted. We have a single molecule in our system, which means that the concentration in our system is infinitely low compared to that
in experiments; in addition our system has zero salt concentration, which might also have an effect on the prefactor.

Our inability to observe a clear transition with temperature for the shortest peptide chain, (GVG)(VPGVG)$_3$ is not surprising. Conflicting results have been reported regarding the existence of an LCST for short ELP chains; these may be because the investigators all considered different system conditions. Nuhn et al. \textsuperscript{42} investigated ELPs of several lengths with the repeating unit (GVGVP) in aqueous NaCl solutions containing 10 mg/mL peptide and did not find an LCST for peptides shorter than or equal to (GVGVP)$_3$. The LCST of (GVGVP)$_4$ determined by extrapolation of the LCSTs at different salt concentration to zero-salt concentration is approximately 151 °C, which is quite close to what we get from extrapolating our fitted power law function for (GVG)(VPGVG)$_3$. In a simulation study on a single chain GVG(VPGVG) \textsuperscript{43}, Rousseau et al. claimed that the LCST they found was approximately 60 °C, a value outside the range of the experimental values. In a much longer simulation on GVG(VPGVG)$_3$ \textsuperscript{45}, Krukau et al. reported that the transition temperature of this sequence to be 310 K based on analysis of Rg and the size distribution of the hydration shell. However, there was no abrupt change in this value as the temperature rose from below to above $T_t$.

The observed hydration behavior of ELPs show that as temperature increases, less H-bonds are formed between peptide and water ($N_{pw}$), while more H-bonds are formed within the peptide itself ($N_{pp}$). This indicates that water tends to avoid the peptide and the peptide tends to interact more with itself as temperature increases.
We found that the changes with temperature in the number of H-bonds per pentamer between peptide and water and within the peptide did not depend on chain length. Therefore, the ability of one pentamer to gain H-bonds between peptide and peptide or to lose H-bonds between peptide and water as temperature increases depends only on the sequence not on the length of the peptide in which it is embedded. We further investigated the loss and gain of H-bonds by a single pentamer when temperature rises from below to above $T_t$, and found that there is close to one H-bond between pentamer and water lost and 0.13 H-bond within peptide gained.

The percentages of the various secondary structures were calculated. The results suggested that the peptide becomes more ordered when temperature rises from below to above $T_t$. Meanwhile, it is harder for the longer chains to be as ordered as the shorter ones based on our observation. Peptides became more hydrophobic upon heating according to our SASA analysis, which is consistent with our previous investigation $^{47}$ for the single chain (VPGVG)$_{18}$. Moreover, the interaction energy analysis indicated that there are less peptide-water interactions and more peptide-peptide interactions as temperature rises from below to above $T_t$.

### 4.5 Conclusion

There are three major findings in this paper. First, the LCST behavior of ELP molecules originates from directional forces, hydrogen bonding; there is no indication that the compressibility effects associated with polymer-like behavior contributes to LCST, even
though some of these chains are rather long (up to 150 amino acids). Second we find strong
evidence for an abrupt change in ELP properties associated with an LCST-type transition
temperature in a single molecule; this is a surprise since LCST behavior of ELPs is
invariably reported for multiple-chain systems and is considered evidence for collective
behavior. While LCST behavior is certainly a collective phenomenon and the Tc do depend
on peptide concentration, our study shows that the seeds for LCST behavior, that is the
competition between water-peptide and peptide-peptide interactions, are evident at the single
molecule level. A clear transition with temperature is observed in the ELPs’ hydration
properties, more specifically the number of H-bonds between peptide and water and the
number of water molecules within the first hydration shell of the peptide backbone. A third
finding is that the tendencies of each pentamer within an ELP to form water-peptide and
peptide-peptide hydrogen bonds as temperature increases is independent of the ELP length.

ASSOCIATED CONTENT

Supporting Information Available. This material is available free of charge via the Internet at
http://pubs.acs.org

Further details of high intensity regions in Ramachandran plots of residues in peptides,
distribution of water hydrogen bonding network, time autocorrelation function of peptide
backbone, radial distribution function of water to peptides, secondary structure formation as a
function of temperature at different time intervals, number of water molecules in the hydration shell versus temperature and hydrogen bonding lifetime are provided (PDF)

AUTHOR INFORMATION

Corresponding Author

*To whom correspondence may be addressed: Dr. Carol K. Hall, Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27606, United States.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. *These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the NSF’s Research Triangle MRSEC (DMR-1121007). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.
4.6 References


Table 4.1 Summarizing properties of ELPs at lengths (L=10, 18 and 30) averaged over temperatures below $T_t$, above $T_t$ and the difference between them.

<table>
<thead>
<tr>
<th>Length</th>
<th>SASA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>E&lt;sub&gt;hp&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>$%\beta_{\text{turn}}$&lt;sup&gt;c&lt;/sup&gt;</th>
<th>$%\beta_{\text{strand}}$&lt;sup&gt;d&lt;/sup&gt;</th>
<th>E&lt;sub&gt;pw&lt;/sub&gt;&lt;sup&gt;e&lt;/sup&gt;</th>
<th>E&lt;sub&gt;pp&lt;/sub&gt;&lt;sup&gt;f&lt;/sup&gt;</th>
<th>N&lt;sub&gt;pp&lt;/sub&gt;&lt;sup&gt;g&lt;/sup&gt;</th>
<th>N&lt;sub&gt;pw&lt;/sub&gt;&lt;sup&gt;h&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
<td>penta</td>
</tr>
<tr>
<td></td>
<td>General properties of ELP below $T_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>345.47</td>
<td>160.98</td>
<td>1.78</td>
<td>0.74</td>
<td>-96.32</td>
<td>-58.42</td>
<td>0.84</td>
<td>6.22</td>
</tr>
<tr>
<td>18</td>
<td>295.69</td>
<td>168.28</td>
<td>1.03</td>
<td>0.47</td>
<td>-85.04</td>
<td>-57.84</td>
<td>1.07</td>
<td>7.37</td>
</tr>
<tr>
<td>30</td>
<td>280.34</td>
<td>180.54</td>
<td>0.67</td>
<td>0.24</td>
<td>-81.82</td>
<td>-59.6</td>
<td>0.81</td>
<td>5.98</td>
</tr>
<tr>
<td></td>
<td>General properties of ELP above $T_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>357.1</td>
<td>157.13</td>
<td>1.79</td>
<td>1.05</td>
<td>-92.83</td>
<td>-56.58</td>
<td>0.82</td>
<td>5.46</td>
</tr>
<tr>
<td>18</td>
<td>264.9</td>
<td>183.53</td>
<td>1.03</td>
<td>0.61</td>
<td>-73.06</td>
<td>-61.84</td>
<td>1.32</td>
<td>6.22</td>
</tr>
<tr>
<td>30</td>
<td>256.51</td>
<td>190.79</td>
<td>0.56</td>
<td>0.35</td>
<td>-70.87</td>
<td>-62.13</td>
<td>0.96</td>
<td>5.11</td>
</tr>
<tr>
<td></td>
<td>$\Delta =$ Properties&lt;sub&gt;above&lt;/sub&gt;-Properties&lt;sub&gt;below&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11.63</td>
<td>-3.85</td>
<td>0.02</td>
<td>0.31</td>
<td>3.49</td>
<td>1.84</td>
<td>-0.02</td>
<td>-0.76</td>
</tr>
<tr>
<td>18</td>
<td>-30.8</td>
<td>15.25</td>
<td>0.03</td>
<td>0.14</td>
<td>11.98</td>
<td>-4.00</td>
<td>0.25</td>
<td>-1.15</td>
</tr>
<tr>
<td>30</td>
<td>-23.82</td>
<td>10.26</td>
<td>-0.11</td>
<td>0.11</td>
<td>10.95</td>
<td>-2.52</td>
<td>0.15</td>
<td>-0.86</td>
</tr>
<tr>
<td>Average $\Delta$</td>
<td>-14.33</td>
<td>7.22</td>
<td>-0.02</td>
<td>0.19</td>
<td>8.81</td>
<td>-1.56</td>
<td>0.12</td>
<td>-0.93</td>
</tr>
</tbody>
</table>
Figure 4.1 Number of hydrogen bonds between peptide and water, $N_{pw}$, versus number of water molecules within the first hydration shell of peptide backbone, $N_w$. (A) (VPGVG)$_{30}$, the points are grouped into two separate clusters suggesting that $T_c$ is between 305 K and 310 K. (B) (VPGVG)$_{18}$ with $T_c$ between 330 K and 335 K. (C) (VPGVG)$_{10}$ with $T_c$ between 340 K and 350 K also has two distinct clusters. (D) GVG(VPGVG)$_3$ does not have distinct clusters. Clusters below $T_c$ are colored red, clusters above $T_c$ are colored blue. The standard deviation is based on the time average of each simulation.
Figure 4.2 Transition temperature versus chain length from experiment (blue squares) and simulation (red triangles) based on Npw-Nw analysis. Both data sets are fitted with power function $T_t = C^*_{t} L^{-\nu}$. The coefficient from experiments is $C_e = 473.81$, and from simulations is $C_s = 354.93$. The exponent from experiments is -0.63, and from simulations is -0.65. The simulation $T_t$ value for $(GVG)(VPGVG)_3$ (blue empty square), $157 \, ^\circ C$, is obtained by extrapolating the fitted power function to the length 3.5.
Figure 4.3 Number of H-bonds between (A) peptide and water, $N_{pw}$, and (B) peptide and peptide, $N_{pp}$, versus temperature. Slopes of linear fit of all the curves are shown next to each curve. Inset graphs depict the average values of the slope per pentamer for different lengths.
Figure 4.4 (A) Difference between the average of $n_{pw}$ or $n_{pp}$ over all the temperatures above $T_t$ and the average of $n_{pw}$ or $n_{pp}$ over all the temperatures below $T_t$ for each pentamer, $\Delta n_{pw} = n_{pw}^{\text{above}} - n_{pw}^{\text{below}}$, $\Delta n_{pp} = n_{pp}^{\text{above}} - n_{pp}^{\text{below}}$. (B) Schematic representation of the configurations of ELP at temperatures (a) below $T_t$ and (b) above $T_t$. The peptide backbone is grey, peptide-water H-bonds are green and peptide-peptide H-bonds are magenta. Water molecules are represented by 3 circles; circles in red are oxygen atoms and in blue are hydrogen atoms.
Chapter 5 The Role of Amino Acid Order in Temperature-responsive Properties of Elastin-like Polypeptides: Molecular Dynamics Simulations of (VPGVG)$_{18}$ and (VGPVG)$_{18}$

5.1 Introduction

Elastin-like peptides (ELP) are a class of synthetic polypeptides with elastin-based sequences, which undergo a solubility transition upon changes in temperature.\(^1\) ELPs, which exhibit LCST behaviors, are soluble in aqueous solutions below their transition temperature $T_t$, but hydrophobically collapse and aggregate at temperatures greater than the $T_t$.\(^2, 3\) ELPs are characterized by a high sequence control, temperature responsiveness and biocompatibility, which make them highly interesting as a class of synthetic polymers for a variety of applications ranging from drug delivery\(^4-6\) to protein purification\(^7-9\) and from responsive nanoparticles\(^10\) to tissue engineering\(^11, 12\). The most widely studied ELP is poly(VPGVG) based on its LCST behavior as well as its responsive behaviors on the changes in concentration\(^13-15\), pH\(^16\) or ionic strength\(^17\) in aqueous solutions. This pentapeptide motif was also generalized into the VPGXG by Urry’s group, where fourth X is often termed the “guest’ residue, as it can be substituted with other amino acids except for proline.\(^14, 18\)

ELPs that exhibit LCST behavior may also display temperature-dependent hysteresis in their phase transition behavior. For example, it was demonstrated that poly(VPAVG) display significant thermal hysteresis in its heating-cooling cycle.\(^19, 20\) A considerable
undercooling was needed to resolubilize poly(VPAVG)s. In a recent study from Chilkoti’s group, the thermo-hysteresis behavior found for poly(VPAGVG) showed similarities but also significant differences to poly(AVPGVG), despite the fact that poly(VPAGVG) and poly(AVPGVG) have the same amino acid composition and molecular weight. The thermo-hysteresis behavior of certain ELPs opens new possibilities for materials that not only exhibit a smart response under different stimuli, but also that can remain in their states after the stimulus has finished.

Generally, a polypeptide sequence may differ from others in the composition and the order of its constituting amino acids. The sequence order may have a major effect in the function and property of polypeptides even with a given composition. Chilkoti’s group synthesized poly(VPGVG) and poly(VGPVG), which are in opposite directions to each other, and studied their thermal responsive properties by means of turbidity measurements. Although both polypeptides have very close $T_t$ (in the interval 58–53 °C), their thermoresponsive behaviors in water are different. After heating above $T_t$, forward sequence poly(VPGVG) forms a voluminous easily sedimenting suspension, which becomes soluble again on cooling below $T_t$. On the contrary, backward sequence poly(VGPVG) shows a strong hysteresis of the thermal solubility transition.

Although turbidity measurements can precisely describe the macroscopic transition, their molecular nature might be very subtle and the detection is not as easy as the detection of macroscopic phase transition. Furthermore, a comparison of poly(VPGVG) with poly(VGPVG) gives a good example of a strong dependency between the arrangement of
amino acid in the local repeated sequence and thermo-responsive behaviors. In our previous study, the effect of temperature on the structure, dynamics and association of (VPGVG)$_{18}$ in aqueous solution is investigated using long time-scale atomistic molecular dynamics simulations. In this work, we further utilized molecular dynamics simulation technique to reveal the significance of sequence order on the thermal behaviors of ELPs (VPGVG)$_{18}$ and (VGPVG)$_{18}$. We aim to explain the mechanism behind different hysteresis behaviors observed in experiments in phase transition processes.

### 5.2 Materials and Methods

In this study, we employed the Amber 11 and 12 suites of programs to run MD simulations. The ff99SB force field was employed to describe the peptides and water was modeled using the standard TIP3P model. The non-bonded interactions were truncated at a 9 Å cutoff along with a 0.00001 tolerance for Ewald convergence and the long-range electrostatic interactions were taken into account by particle mesh Ewald (PME) summation.

Single–peptide simulations were performed at ten temperatures between 290 K and 350 K to span the temperature range which is relevant to the experimental study of LCST phase transitions of poly(VPGVG) and poly(VGPVG) in aqueous solutions. The initial structures for the models of elastin-like polypeptides are built based on dihedral angles for each residue in Urry’s β-spiral model, which were then solvated in explicit water with the closest distance between any polypeptide atom and the edge of the periodic box to be at least
8 Å to ensure that the polypeptide doesn’t cross periodic boundaries to interact with its own image. The simulation box for \((\text{VPGVG})_{18}\) contained 6856 water molecules with the size of about 5.9 nm × 7.7 nm × 5.0 nm in the equilibrium state and the simulation box for \((\text{VGPVG})_{18}\) contained around 8651 water molecules with the size of 6.9 nm × 8.7 nm × 5.7 nm in equilibrium state. The equilibrium box size may slightly vary at different temperatures. The system is equilibrated in eight stages, including several minimization, thermalization and MD run cycles. The equilibration protocol was the same as described in our previous study\(^{22}\). The system temperature was maintained using a Berendsen thermostat.\(^{28}\) The SHAKE algorithm was used to constrain the position of the hydrogen atoms.\(^{29}\)

The production simulations were performed for at least 70 ns with a 2 fs time step. To ensure the convergence of our simulations, we calculated the time autocorrelation function for in-plane and out-of-plane backbone rotations, from which the relaxation time of peptide backbone was estimated to be less than 25 ns (Figure 5.S3, S4). The statistical and clustering analysis was carried out for the last 40 ns MD trajectories using in-house scripts along with the standard tool suite accompanying Amber12.0. The interaction energy was calculated using the molecular mechanics energy function in NAMD 2.\(^{30}\) The hydrogen bond analysis was performed using an angle cut-off of 30° and a distance cutoff of 3.5 Å.

The trajectories from the single-peptide MD simulations were clustered to produce three structural clusters using the pairwise RMSD between frames as a metric comparing the atoms named CA with a critical distance of 12 Å. This clustering analysis was conducted based on a hierarchical RMSD-based clustering algorithm.\(^{31}\) Representative structures from
the most populated clusters at 290 K and 350 K are shown in Figure 5.S1 and 5.S2. The lowest-energy representative single polypeptide structures were then taken to be the initial structures in the simulations of the interaction between double polypeptides. The two single polypeptide structures chosen were placed side-by-side with a 32.3~32.4 Å distance between the centers of mass of the polypeptides and a 7 Å distance between the two closest surfaces. The structures were then solvated in explicit solvent with the closest distance between any solute atom and the edge of the periodic box to be 12 Å and simulated at 350 K.

For simulations with multiple peptides, the simulation box contained 10493 water molecules with the size of 9.2 nm×7.9 nm×5.8 nm for double-poly(VPGVG) system and 9738 water molecules with size of  6.6 nm×9.4 nm×6.4 nm for double-poly(VGPVG) system. Equilibration protocols and MD simulations were the same as described for single-peptide simulations. The production simulations were carried out for at least 15 ns with a 2 fs time step.

5.3 Results

5.3.1 Structural properties and water-peptide interactions

In this study, the temperature-dependent behaviors of forward sequence (VPGVG) 18 and backward sequence (VGPVG) 18 in aqueous solutions have been studied by atomics molecular dynamics simulations. The temporal evolution of $R_g$ at various temperatures ranging from 290 K to 350 K for poly(VPGVG) and poly(VGPVG) systems are displayed in Figure 5.S1. As discussed in our previous paper, the single (VPGVG) 18 chain adopts a
collapsed state over a temperature range between 290 K and 350 K. The same phenomena was observed for the backward sequence (VGPVG)\textsubscript{18} (Figure 5.1, S.1). The averaged value of radius of gyration of (VPGVG)\textsubscript{18} declines slightly as temperature increases. However, for backward sequence (VGPVG)\textsubscript{18}, the R\textsubscript{g} stays approximately the same at high and low temperatures.

We then compared the number of intrachain hydrogen bonds with (n\textsubscript{wp}) and without (n\textsubscript{pp}) mediated water in two systems (Figure S.5a). Unlike the monotonically decreasing/increasing trend observed in (VPGVG)\textsubscript{18} system, both n\textsubscript{wp} and n\textsubscript{pp} only slightly vary upon heating up from 290 K to 350 K in (VGPVG)\textsubscript{18} system, which correlates with the stability of radius of gyration of single (VGPVG)\textsubscript{18} chain.

We further examined the effect of amino acid order of the repeated unit on the structure and hydrophobicity of ELPs by characterizing the peptide-water interaction properties. For (VPGVG)\textsubscript{18}, solvent accessible surface area (SASA) and number of water-peptide hydrogen bonds (n\textsubscript{pw}) slightly decrease upon heating up to ~330 K and then experience a sudden decrease up to 350 K. In contrary, The dependence of SASA and n\textsubscript{pw} on the temperature for (VGPVG)\textsubscript{18} showed a quite different pattern. At high temperature (VGPVG)\textsubscript{18} has higher solvent accessible surface area (SASA) and forms more hydrogen bonds with water molecules than (VPGVG)\textsubscript{18} peptide. This fact clearly reflects the reverse in the amino acid order makes the peptide surface more hydrophilic at elevated temperatures.

The number of water molecules in the first hydration shell is defined as the number of water molecules within a 2.23 Å distance of any atom on the peptide. This distance is taken
to be the distance at which the radial distribution function (RDF) between the oxygen of water and an atom on the peptide has its first minimum.\textsuperscript{22} As shown in Figure 5.S5b, there are negligible differences in the number of first layer hydration water around \((\text{VPGVG})_{18}\) and \((\text{VGPVG})_{18}\) at the temperature below 330 K. At high temperature, however, the differences become more significant. There is a sudden decrease in the \((\text{VPGVG})_{18}\) system which is corresponding to the sudden decrease in the SASA profile.

The interaction energy between both polypeptides and water increases as temperature increased because the peptide and water interaction becomes energetically unfavored as the temperature increases (Figure 5.1d). We observed smaller interaction energy values for backward sequence \((\text{VGPVG})_{18}\) than forward sequence \((\text{VPGVG})_{18}\), which indicates the surface hydrophobicity of single \((\text{VGPVG})_{18}\) is smaller than \((\text{VPGVG})_{18}\) at high temperatures.

Overall, the amino acid order reversing appears to cause minor structural and energetic changes in ELPs at low temperature. However, the poly(\text{VGPVG}) is generally more hydrophilic than poly(\text{VPGVG}) at high temperatures, although their chemical structures are identical.

5.3.2 Dihedral angles, secondary structures and role of individual residue

In order to explore the structural motifs which lead to their different temperature-responsive behaviors in aqueous solutions, we calculated the torsion angles of each residue in VPGVG and VGPVG pentamers in \((\text{VPGVG})_{18}\) and \((\text{VGPVG})_{18}\), respectively. The Ramachandran plots are displayed in Figure 5.2 and 5.S6. The high intensity region in
Ramachandran plot indicates the most favorable, low-energy torsion angle pairs. In our previous study, we discussed the features of torsion angles in VPGVG pentamer at low and high temperatures and compared our simulation results with multiple experimental studies in exhaustive details. The situation found for backward sequence (VGPVG) showed similarities but also significant differences to forward sequence (VPGVG), especially for valine residues (Figure 5.2).

Val1 in VPGVG is preceded proline which shows distinctive torsion angle distribution from the other valines in both VPGVG and VGPVG pentamer. The ψ angle of Val1 in VPGVG only populate around 150° and φ is around −130° or −75°, whereas Val4’ torsion angles are distributed between four regions where ψ is around 150° or −2° and φ is around −130° or −90°. However, Val1 and Val4 in VGPVG occupy similar conformational spaces in the left side of the Ramachandran plots. The ψ angle display the bimodal distribution with (ψ ~150° or −5°, respectively) while φ ranging from ~−45° to ~−170° with one peak around −126° for a4_1, a4_2, and d4_1, and −107.7° for d4_2. The torsion angles of Val4 residues in VGPVG also clustered at the left-handed alpha helical conformational space marked as d4_3 (52°, 30.3°).

Glycine and proline are usually observed to have very different ϕ, ψ distribution outlines than the other amino acids, with conformational constraints either significantly less or significantly more. In VPGVG, the Pro2 and Gly3 primarily adopt the torsion angle pairs (ϕ = −60°, ψ = 120°), (ϕ = −75°, ψ = 150°) as the i+1 and i+2 residues of a type-II β-turn structure. Gly2 in VGPVG is also a pre-proline residue; however, Gly2 does not have a
side chain which allows high flexibility as well as the torsion angles which are normally not allowed for other amino acid residues. Moreover, the most populated region for Gly3 in VPGVG is missing in the plots for Gly2 in VGPVG, indicating the omitting of turn structures in the segment VGPV. The most populated angles adopted by Gly2 in VGPVG correspond to the high intensity regions of the Ramachandran plot labeled by c3_1 (−76.5°, 150.6°), c3_2 (70°, −141.1°) and c4_1 (98.27°, −140.0°) as shown in Figure S6. The Gly5 in VPGVG are scattered over various conformational spaces of the Ramachandran plot, which may adopt various secondary structural motifs such as Helix, Turn and β-strand. Interestingly, the Ramachandran plot pattern for Gly5 in poly(VGPVG) at 350 K is similar to Gly5 in poly(VPGVG) at 290 K; the Ramachandran plot pattern for Gly5 in poly(VPGVG) at 290K is similar to Gly5 in poly(VPGVG) at 350 K. Gly5 residue in poly(VGPVG) at 290 K and 350 K occupy two areas with relatively high number of occurrences, such as e3_1 (−80.76°, 135°) and e4_2 (−107.7°, −2.0°), respectively.

In contrast to glycines, ψ angle for Pro3 in VGPVG varies in a wider range from −62° to 173° which indicates a larger flexibility of proline residues in VGPVG than that in VPGVG, where Φ is fixed around −65° due to its backbone stereochemistry. The representative structures for pentamer VPGVG and VGPVG are built based on the most populated torsion angles obtained from Ramachandran plots, as shown in Figure 5.3e and f.

The secondary structural motifs are determined by a widely used structure recognition algorithm DSSP method, which is based on H-bonding patterns. Along with the DSSP program, the residues bracketed by the hydrogen bond (i, i+n (n=3, 4, 5)) are considered as to
adopt turn structure. The helical-structure includes residues from \( \alpha \), 3-10 and \( \pi \) helix; the \( \beta \)-strand structure includes residues from \( \beta \)-bridges and extended strands involved in parallel and anti-parallel \( \beta \)-sheets. The secondary structural propensities of for each single polypeptide at various temperatures are listed in Table 5.S2. Distinct changes were observed due to the reversing of the amino acid order in the secondary structural propensities. In general, there is a larger propensity for helix structures in backward sequence \((\text{VGPVG})_{18}\) chain than that in the forward sequence \((\text{VPGVG})_{18}\) at all temperatures, which is consistent with the observation in the Ramachandran plot. In addition, \((\text{VPGVG})_{18}\) systems adopt more Turn structures than \((\text{VGPVG})_{18}\) for most cases. The propensities for \( \beta \)-strands are very close for two polypeptides at low temperature; however, there is a smaller propensities for \((\text{VGPVG})_{18}\) chain at high temperature than \((\text{VPGVG})_{18}\), which is probably due to a higher flexibility of prolines and valines in poly(VGPVG). The occurrence of each residue in each type of secondary structural motifs (Turn, \( \beta \)-strands, and Helix) at 290 K and 350 K is presented in Table 5.S3 and plotted in Figure 5.S7.

In forward sequence VPGVG, Pro2-Gly3-Val4 segment mostly occurred in Turns and Gly5-Val1 mostly occurred in \( \beta \)-strands. In backward sequence VGPVG, residues adopt diverse secondary structures. Specially, more than 8% of each residue in Gly2-Pro3-Val4 segment adopt Helix structures. Aforementioned, proline has its own asymmetrical structure and restrains the residues before it, which can break the symmetry of peptide chains. The secondary structures differ significantly due to such kind of effect, which demonstrate the
importance of the direction of residues and the local amino acid order on the secondary structures of peptides.

Because any pair interaction can be affected by the order of two interacting residues, the reversing of amino acid order from poly(VPGVG) to poly(VGPVG) can be described at the molecular level to be the mutation of Pro2-Gly3 to Gly2-Pro3. Statistical analysis was conducted on the distances (D1, D2) between the carbon α atoms in Val1 and Val4 in both forward and backward sequences based on the last 40 ns trajectories. As listed in Table 5.1, 11.1% and 33.3% of Val1 and Val4 pairs form hydrogen bonds (D1< 7 Å) in forward sequence at 290 K and 350 K, respectively. Interestingly, 33.3% and 11.1% of Val1 and Val4 pairs form hydrogen bonds (D2< 7 Å) in the backward sequence at 290 K and 350 K, respectively. The difference is due to the fact that the segment VPGV intrinsically adopt type-II β-turn structures which are stabilized by hydrogen bonds between Val1 and Val4 and the number of β-turn structures increased as the temperature is increased from 290 K and 350 K. Therefore, the mutation of Pro2-Gly3 to Gly2-Pro3 cause significant differences in the local residue interactions. The results further addressed the effect of amino acid order on secondary structures measured above.

Interactions between each amino acid residue with their aqueous environment, determine the conformation of polypeptide and mediate further intermolecular interactions. Insight into the role of amino acid in their thermal-responsive behaviors was obtained by means of residue-water interaction energy, SASA, and residue-water hydrogen bonding. The interaction energies of each amino acid residue in the repeat unit of VPGVG and VGPVG in
(VPGVG)$_{18}$ and (VGPVG)$_{18}$, respectively, with water molecules were calculated (Figure 5.3a). In general, the interactions between each amino acid and water molecules weakens as the temperature is increased, except for Val4 in backward sequence VGPVG. The energy plots clearly confirm that the interaction between each residue and water in the two sequences is not identical, although Pro2/3 and Val4-Gly5 in these two sequences can be argued to show the very similar hydrophobicity. Val1 in the VPGVG interacts with water more weakly than Val1 in VGPVG, possibly due to the steric restriction of Val1 in VPGVG by Pro2.

The number of hydrogen bonds formed between each residue and water molecules are correlated with the interaction between each residue and water which reveals a less structural heterogeneity in VGPVG than VPGVG. (Figure 5.3b) The SASA of each residue in two sequences was also shown in Figure 3c. The SASA not only provide a way to define the peptide surface and interior, but also has physical significance as a gain in hydrophobic interaction free energy is accompanied by a reduction in the SASA.$^{34}$ The plot of SASA for each residue reveals very little difference between the hydrophobicity of Pro2, Val4 and Gly5 in VPGVG and Pro3, Val4 and Gly5 in VGPVG. However, the Val1 in VGPVG is more accessible by water molecules than Val1 in VPGVG and Gly2 in VGPVG is less accessible than Gly3 in VPGVG. The compensation between the SASA of these two residues determines the quantitative relationship between the overall SASA of both ELP chains at different temperatures.
5.3.3 Hydration water network

The thermal-responsive behaviors of ELPs not only involve changes in the peptide conformation and water-peptide interaction, but also involve the changes in water structures around the peptide. In other words, not only water influences peptide properties but the peptides as well can modify water structure and dynamics.\(^{35}\) We further characterize the structure of hydration water around the peptide chains by measuring the size of the largest H-bonded water network around the polypeptide (Figure 5.4a), which represents the effect of orientational ordering of water molecules and water-water interactions near peptide surface.

The characterization details are listed in our previous paper.\(^{22}\) In the hydration water network, all water molecules are connected by hydrogen bonds with \(R_{\text{OO}} \leq 3.5\ \text{Å}\) and \(\varphi_{\text{O1-O2-H2}} \leq 30^\circ\) (Figure 5.4b). The last 40 ns trajectories were examined to identify the largest network of water molecules at each temperature. Probability distribution of the largest water network around the (VPGVG)\(_{18}\) and (VGPVG)\(_{18}\) at \(T = 290\ \text{K}, 310\ \text{K}, 330\ \text{K}\) and \(350\ \text{K}\) are displayed in Figure 5.4a.

For both peptides, the WAT-WAT H-bonds in the hydration water generally decreases with increasing temperature, indicating a greater disordered water network as temperature increase, especially for (VGPVG)\(_{18}\) chain. A larger water network at low temperature was considered to limit the freedom of movement of the peptide, whereas at high temperature, the peptide gain flexibility and explores further conformation or intermolecular interactions more easily. We noticed a higher thermal stability of the H-bonded water network around (VPGVG)\(_{18}\) than that around (VGPVG)\(_{18}\), whereas the hydration water of
(VGPVG)$_{18}$ has a higher thermal-sensitivity than that of (VPGVG)$_{18}$. Aforementioned, there are more hydration water molecules in the watershell of (VGPVG)$_{18}$ which also has a higher SASA; however, the water molecules form more hydrogen bonds with (VGPVG)$_{18}$, which may damage the water-water connectivity in the hydration shell. The trade-off between this two effects determines the structure of hydration water around ELPs.

5.3.4 Double polypeptide aggregation

To examine the effect of temperature on the early stage of peptide aggregation between two ELP chains, we performed simulations of two temperature-equilibrated polypeptides at 350 K. The simulation set-up and MD simulation protocols were the same as described in our previous paper. $^{22}$ Several interaction parameters were calculated, namely, distance between the center-of-mass of two chains ($D_{cc}$), interaction energy ($E_{pp}$), and number of interchain hydrogen bonds ($N_{pp}$). During the simulation trajectories, in both systems, two polypeptides assemble into aggregates with a certain degree of compactness (Figure 5.5). Polypeptides with backward sequence VGPVG form more compacted aggregates with more inter-chain hydrogen bonds formed and a lower interaction energy between two chains.

Previously, the clear hysteresis in the heating and cooling cycle was also observed in the PNIPAM water solution. $^{36-38}$ It was demonstrated that some of interchain hydrogen bonds could persist in the cooling, especially when the temperature is not far away from the phase transition temperature by a FTIR spectra. $^{39}$ Furthermore, a comparison between PNIPAM and PDEAM indicates that the hysteresis originates from some additional hydrogen
bonds formed in their collapsed state at temperatures higher than the LCST. The hysteresis behavior of poly(VGPVG) is probably because of a strong interaction in the aggregate of poly(VGPVG) chains with more interchain hydrogen bonds formed. These interchain hydrogen bonds can stabilize the aggregated phase by providing stable interchain contacts that are not easily disrupted by cooling. Aforementioned, (VGPVG)$_{18}$ adopt a more extended conformation and larger surface area in (VGPVG)$_{18}$ are exposed, which may increase the chances for the formation of intermolecular hydrogen bonds.

5.4 Conclusion and Discussion

This paper presents the results from a comprehensive atomistic molecular dynamics study on the effect of temperature and amino acid order on the structural and dynamics changes of the ELP chains and their hydration properties. This work was motivated by the different thermos-hysteresis behaviors of forward sequence poly(VPGVG) and backward sequence poly(VGPVG) peptides detected in the heating-cooling cycle of their aqueous solutions: poly(VPGVG)s showed a direct dissociation upon cooling below Tt; however, strong undercooling was needed to resolubilize poly(VPGVG)s.

We estimated the role of the amino acid order in peptides for conformational properties, hydrophobicity, local residue interactions, and the secondary structure propensity in a range of temperatures. (VPGVG)$_{18}$ and (VGPVG)$_{18}$ both show an increased hydrophobicity at elevated temperatures in terms of water-peptide hydrogen bonds, SASA, and interaction energy between peptide and water molecules, which is the common
characteristic feature of neutral thermos-responsive polymers showing the LCST behavior. However, (VGPVG)$_{18}$ is more hydrophilic than (VGPVG)$_{18}$ at high temperatures despite the fact that they have the same amino acid composition and molecular weight. (VPGVG)$_{18}$ also assumes a more compacted structure as indicated by a smaller radius of gyration at high temperature.

Our simulations show that at high temperature, (VPGVG)$_{18}$ tend to form more ordered secondary structures in the form of β-strands and β-turns than (VGPVG)$_{18}$. However, the changes in temperatures for (VGPVG)$_{18}$ do not cause the notable changes of the secondary structures. At all temperatures, the (VGPVG)$_{18}$ contains more helices than (VPGVG)$_{18}$ and a smaller amount of β-strands and β-turns than (VPGVG)$_{18}$ at high temperatures. Such a dependency between amino acid order and secondary structures was reported before. It was reported that through the examination of 1288 PDB structural fragments, about 21% of the protein sequence segments exhibits different secondary structure propensities in forward and backward directions assessed by PREDATOR algorithm.$^{41,42}$

The “Pro-Gly” sequence motif was demonstrated to be important for the elastomeric mechanical behaviors observed within a number of native protein-based polymers.$^{43}$ It is reasonable to assume that it also a recurrent if not dominant structural features associated with the reversibility in the LCST behaviors of ELPs. This is because the type II β-turn per pentamer VPGVG involves proline-glycine pair as the corner residues. The presence of hydrogen bond between Val1 and Val4 in VPGV is important to stabilize the type-II β-turn structure. The switch of Pro-Gly into Gly-Pro still leads to a functional polymer which
exhibited LCST behaviors, but destroy the β-turns which resulting a more extended conformation (larger \( R_g \) and smaller \( n_{pp} \)) and a larger solvent exposure (larger SASA).

We also examined the association kinetics of poly(VPGVG) and poly(VGPVG) in water in the double peptide simulations, which revealed more details in interchain association. In analogy to poly(VPGVG), the association of two poly(VGPVG) above LCST can also be regarded as the consequence of the competition between hydrophilic polymer–water interactions and hydrophobic polymer–polymer interactions. The tracked structural and energetic characteristics include distance between the center-of-mass of two peptides, interaction energy between peptides and number of inter-chain hydrogen bonds. Poly(VGPVG)s form stronger aggregation with more a larger number of inter-chain hydrogen bonds, which is probably due to a more extended conformation and larger solvent exposure of single poly(VGPVG). Simulation results indicate that the hysteresis may originate from some additional interchain hydrogen bonds.

The distinct coacervation process, which is exhibited by ELPs upon heating, is considered to be a complex and multistep transition in thermodynamics point of view.\(^{20, 44, 45}\) It was proposed that the coacervation process can be simplified into intra-molecular interaction and inter-molecular aggregation stages, and intra-molecular interactions during the initial chain folding prevail.\(^ {46, 47}\) Thus, it is reasonable to assume that the dissociation in the cooling process would occur in the reverse order of the two interaction stages; \( i.e. \) the order of intrachain interaction and interchain interaction is reversed. Upon cooling, the dissociation of the inter-chain aggregation occurs first. Collectively, the consecutive inter-
chain hydrogen bonding arising in the poly(VGPVG) appeared to perturb the dissociation by the additional interchain interactions. In other words, each poly(VPGVG) aggregate directly and quickly dissociates back into individual chain when the solution is cooled to a temperature below the LCST, which is different from the dissociation of poly(VGPVG) aggregates in which some aggregates are not able to completely dissociate due to some extra inter-chain hydrogen bonds formed.

Based on our simulations, we conclude that the conformational preferences of the constituent amino acids in a specific order play an important role in the structure, dynamics and thermodynamics properties of ELPs (VPGVG)$_{18}$ and (VGPVG)$_{18}$.

Funding Sources

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107).

ACKNOWLEDGMENT

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107). The computer support was provided by the High Performance Computing Center at North Carolina State University.
5.5 References


Table 5.1 Distance between Cα@Val1-Cα@Val4 in pentamer (VPGVG) and (VGPVG) at 290 K and 350 K, respectively.

<table>
<thead>
<tr>
<th></th>
<th>&lt;7Å</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>(VPGVG)18, 290 K</td>
<td>11.11%</td>
<td>8.29±1.24</td>
</tr>
<tr>
<td>(VPGVG)18, 350 K</td>
<td>33.33%</td>
<td>8.41±1.49</td>
</tr>
<tr>
<td>(VGPVG)18, 290 K</td>
<td>33.33%</td>
<td>7.53±1.22</td>
</tr>
<tr>
<td>(VGPVG)18, 350 K</td>
<td>11.11%</td>
<td>8.60±1.33</td>
</tr>
</tbody>
</table>
Table 5.S1  Dihedral angles obtained from the highest intensity points in Ramachandran plots in Figure 5.S6.

<table>
<thead>
<tr>
<th>VPGVG</th>
<th>$\phi, \psi$</th>
<th>VGPVG</th>
<th>$\phi, \psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1_1</td>
<td>$-131.7^\circ, 156.0^\circ$</td>
<td>a3_1</td>
<td>$-135.4^\circ, 137.0^\circ$</td>
</tr>
<tr>
<td>a1_2</td>
<td>$-82.0^\circ, 146.0^\circ$</td>
<td>a3_2</td>
<td>$-127.2^\circ, -14.17^\circ$</td>
</tr>
<tr>
<td>a2_1</td>
<td>$-131.7^\circ, 150.0^\circ$</td>
<td>c3_1</td>
<td>$-76.5^\circ, 150.6^\circ$</td>
</tr>
<tr>
<td>a2_2</td>
<td>$-70.0^\circ, 140.0^\circ$</td>
<td>c3_2</td>
<td>$70^\circ, -141.1^\circ$</td>
</tr>
<tr>
<td>b1_1</td>
<td>$-61.4^\circ, 150.6^\circ$</td>
<td>b3_1</td>
<td>$-68.0^\circ, -13.5^\circ$</td>
</tr>
<tr>
<td>b2_1</td>
<td>$-62.4^\circ, 150.6^\circ$</td>
<td>d3_1</td>
<td>$-131.5^\circ, 147.0^\circ$</td>
</tr>
<tr>
<td>b2_2</td>
<td>$-65.4^\circ, -23.02^\circ$</td>
<td>d3_2</td>
<td>$-126.1^\circ, -3.5^\circ$</td>
</tr>
<tr>
<td>c1_1</td>
<td>$82.9^\circ, 1.9^\circ$</td>
<td>e3_1</td>
<td>$-80.76^\circ, 135^\circ$</td>
</tr>
<tr>
<td>c1_2</td>
<td>$82.9^\circ, -9.9^\circ$</td>
<td>a4_1</td>
<td>$-126.3^\circ, 147.8^\circ$</td>
</tr>
<tr>
<td>c2_2</td>
<td>$-70.15^\circ, -9.2^\circ$</td>
<td>a4_2</td>
<td>$-127.4^\circ, -7.5^\circ$</td>
</tr>
<tr>
<td>d1_1</td>
<td>$-131.5^\circ, 158.4^\circ$</td>
<td>e4_1</td>
<td>$98.27^\circ, -140.0^\circ$</td>
</tr>
<tr>
<td>d1_2</td>
<td>$-125.8^\circ, 0.024^\circ$</td>
<td>b4_1</td>
<td>$-62.03^\circ, 148.9^\circ$</td>
</tr>
<tr>
<td>d1_3</td>
<td>$-91.68^\circ, 144.5^\circ$</td>
<td>d4_1</td>
<td>$-125.8^\circ, 153^\circ$</td>
</tr>
<tr>
<td>d1_4</td>
<td>$-90.21^\circ, -3.53^\circ$</td>
<td>d4_2</td>
<td>$-107.7^\circ, -2.0^\circ$</td>
</tr>
<tr>
<td>d2_1</td>
<td>$-129.5^\circ, 151.4^\circ$</td>
<td>d4_3</td>
<td>$52.0^\circ, 30.3^\circ$</td>
</tr>
<tr>
<td>d2_2</td>
<td>$-129.5^\circ, -6.9^\circ$</td>
<td>e4_1</td>
<td>$-75.65^\circ, 143.8^\circ$</td>
</tr>
<tr>
<td>d2_3</td>
<td>$-73.89^\circ, 140.5^\circ$</td>
<td>e4_2</td>
<td>$85.74^\circ, -1.5^\circ$</td>
</tr>
<tr>
<td>d2_4</td>
<td>$-79.4^\circ, -7.9^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e1_1</td>
<td>$-78.65^\circ, 150.8^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e1_2</td>
<td>$80.74^\circ, -2.98^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e2_1</td>
<td>$-87.53^\circ, 148.0^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e2_2</td>
<td>$80.74^\circ, -2.98^\circ$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.S2 (a) Secondary structure propensities for (VPGVG)$_{18}$ from simulation at various temperatures.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Turn</th>
<th>Beta–strand</th>
<th>helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>290</td>
<td>0.17638±</td>
<td>0.08482±</td>
<td>1.5E−4±</td>
</tr>
<tr>
<td></td>
<td>0.03252</td>
<td>0.02538</td>
<td>0.00224</td>
</tr>
<tr>
<td>300</td>
<td>0.2201±</td>
<td>0.09539±</td>
<td>0.01029±</td>
</tr>
<tr>
<td></td>
<td>0.04078</td>
<td>0.03591</td>
<td>0.0173</td>
</tr>
<tr>
<td>305</td>
<td>0.22569±</td>
<td>0.04392±</td>
<td>0.00126±</td>
</tr>
<tr>
<td></td>
<td>0.04083</td>
<td>0.02047</td>
<td>0.00644</td>
</tr>
<tr>
<td>310</td>
<td>0.1817±</td>
<td>0.06887±</td>
<td>0.00131±</td>
</tr>
<tr>
<td></td>
<td>0.03798</td>
<td>0.03764</td>
<td>0.00644</td>
</tr>
<tr>
<td>320</td>
<td>0.14165±</td>
<td>0.11286±</td>
<td>0.00674±</td>
</tr>
<tr>
<td></td>
<td>0.03269</td>
<td>0.03389</td>
<td>0.01335</td>
</tr>
<tr>
<td>325</td>
<td>0.11541±</td>
<td>0.07481±</td>
<td>0.0025±</td>
</tr>
<tr>
<td></td>
<td>0.03495</td>
<td>0.02254</td>
<td>0.00876</td>
</tr>
<tr>
<td>330</td>
<td>0.19667±</td>
<td>0.10991±</td>
<td>3.59E−4±</td>
</tr>
<tr>
<td></td>
<td>0.03931</td>
<td>0.03158</td>
<td>0.00346</td>
</tr>
<tr>
<td>335</td>
<td>0.13161±</td>
<td>0.11432±</td>
<td>0.00326±</td>
</tr>
<tr>
<td></td>
<td>0.04283</td>
<td>0.02798</td>
<td>0.01008</td>
</tr>
<tr>
<td>340</td>
<td>0.20706±</td>
<td>0.0746±</td>
<td>7.6E−4±</td>
</tr>
<tr>
<td></td>
<td>0.04345</td>
<td>0.03431</td>
<td>0.00495</td>
</tr>
<tr>
<td>350</td>
<td>0.21557±</td>
<td>0.14108±</td>
<td>0.00758±</td>
</tr>
<tr>
<td></td>
<td>0.03797</td>
<td>0.02946</td>
<td>0.01491</td>
</tr>
</tbody>
</table>
Table 5.S2 (b) Secondary structure propensities for (VGPVG)$_{18}$ from simulation at various temperatures.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Turn</th>
<th>Beta–strand</th>
<th>helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>290</td>
<td>0.12592±</td>
<td>0.07655±</td>
<td>0.03789±</td>
</tr>
<tr>
<td></td>
<td>0.04537</td>
<td>0.0268</td>
<td>0.02674</td>
</tr>
<tr>
<td>300</td>
<td>0.1151±</td>
<td>0.08912±</td>
<td>0.05335±</td>
</tr>
<tr>
<td></td>
<td>0.03251</td>
<td>0.02755</td>
<td>0.02804</td>
</tr>
<tr>
<td>305</td>
<td>0.10554±</td>
<td>0.04973±</td>
<td>0.06663±</td>
</tr>
<tr>
<td></td>
<td>0.03947</td>
<td>0.02119</td>
<td>0.02688</td>
</tr>
<tr>
<td>310</td>
<td>0.09645±</td>
<td>0.10412±</td>
<td>0.09441±</td>
</tr>
<tr>
<td></td>
<td>0.03224</td>
<td>0.02925</td>
<td>0.03254</td>
</tr>
<tr>
<td>320</td>
<td>0.11082±</td>
<td>0.11717±</td>
<td>0.02156±</td>
</tr>
<tr>
<td></td>
<td>0.03886</td>
<td>0.02804</td>
<td>0.02255</td>
</tr>
<tr>
<td>325</td>
<td>0.13925±</td>
<td>0.05094±</td>
<td>0.13217±</td>
</tr>
<tr>
<td></td>
<td>0.04865</td>
<td>0.02576</td>
<td>0.04355</td>
</tr>
<tr>
<td>330</td>
<td>0.09598±</td>
<td>0.07915±</td>
<td>0.05557±</td>
</tr>
<tr>
<td></td>
<td>0.04294</td>
<td>0.02805</td>
<td>0.02458</td>
</tr>
<tr>
<td>335</td>
<td>0.15895±</td>
<td>0.05064±</td>
<td>0.0935±</td>
</tr>
<tr>
<td></td>
<td>0.04435</td>
<td>0.03309</td>
<td>0.04524</td>
</tr>
<tr>
<td>340</td>
<td>0.11873±</td>
<td>0.14028±</td>
<td>0.0341±</td>
</tr>
<tr>
<td></td>
<td>0.04041</td>
<td>0.04684</td>
<td>0.03171</td>
</tr>
<tr>
<td>350</td>
<td>0.13516±</td>
<td>0.06004±</td>
<td>0.05279±</td>
</tr>
<tr>
<td></td>
<td>0.05335</td>
<td>0.03036</td>
<td>0.0334</td>
</tr>
</tbody>
</table>
Table 5.3 (a) The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VPGVG at 290 K.

<table>
<thead>
<tr>
<th></th>
<th>Val1</th>
<th>Pro2</th>
<th>Gly3</th>
<th>Val4</th>
<th>Gly5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>0.01</td>
<td>26.74</td>
<td>37.46</td>
<td>22.54</td>
<td>3.48</td>
</tr>
<tr>
<td>Beta–strand</td>
<td>6.60</td>
<td>14.45</td>
<td>1.64</td>
<td>1.91</td>
<td>14.55</td>
</tr>
<tr>
<td>Helix</td>
<td>1.57e–04</td>
<td>0.20</td>
<td>0.22</td>
<td>0.22</td>
<td>0.11</td>
</tr>
</tbody>
</table>

(b) The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VPGVG at 350 K.

<table>
<thead>
<tr>
<th></th>
<th>Val1</th>
<th>Pro2</th>
<th>Gly3</th>
<th>Val4</th>
<th>Gly5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>0.16</td>
<td>33.99</td>
<td>44.07</td>
<td>28.89</td>
<td>2.95</td>
</tr>
<tr>
<td>Beta–strand</td>
<td>15.51</td>
<td>12.03</td>
<td>8.08</td>
<td>13.17</td>
<td>17.91</td>
</tr>
<tr>
<td>Helix</td>
<td>3.43e–04</td>
<td>0.98</td>
<td>1.07</td>
<td>1.07</td>
<td>0.09</td>
</tr>
</tbody>
</table>

(c) The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VGPVG at 290 K.

<table>
<thead>
<tr>
<th></th>
<th>Val1</th>
<th>Gly2</th>
<th>Pro3</th>
<th>Val4</th>
<th>Gly5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>5.71</td>
<td>10.37</td>
<td>22.56</td>
<td>18.05</td>
<td>6.34</td>
</tr>
<tr>
<td>Beta–strand</td>
<td>10.83</td>
<td>3.13</td>
<td>0.01</td>
<td>8.06</td>
<td>14.59</td>
</tr>
<tr>
<td>Helix</td>
<td>0.05</td>
<td>8.19</td>
<td>8.34</td>
<td>8.34</td>
<td>0.24</td>
</tr>
</tbody>
</table>

(d) The averaged individual propensity on residue occurrences in different types of secondary structural motifs for VGPVG at 350 K.

<table>
<thead>
<tr>
<th></th>
<th>Val1</th>
<th>Gly2</th>
<th>Pro3</th>
<th>Val4</th>
<th>Gly5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>16.55</td>
<td>3.25</td>
<td>13.66</td>
<td>21.65</td>
<td>17.93</td>
</tr>
<tr>
<td>Beta–strand</td>
<td>8.91</td>
<td>3.56</td>
<td>3.29</td>
<td>5.42</td>
<td>3.58</td>
</tr>
<tr>
<td>Helix</td>
<td>1.26</td>
<td>7.99</td>
<td>9.80</td>
<td>10.55</td>
<td>4.30</td>
</tr>
</tbody>
</table>
Figure 5.1 Temperature dependence of properties of a single ELP chain: (a) radius of gyration \( R_g \) of \( (VPGVG)_{18} \) and \( (VGPVG)_{18} \), (b) \( n_{pw} \), the number of peptide-water hydrogen bonds, (c) solvent accessible surface area (SASA). (d) Peptide-water interaction energy. The error bars in these plots represent the standard deviation. The same representations are used in the following figures.
Figure 5.2 (a-d) Ramachandran plot (ϕ, ψ distributions) for Val1 and Val4 in (VPGVG)18 and (VGPVG)18, respectively, at 350 K, colored by intensities. The most populated regions on the Ramachandran plots are labelled. (e-f) Representative structures for G(VPGVG)V segment at 350 K (e) and G(VGPVG)V segment at 350 K (b) built with the dihedral angles obtained from the highest intensity points in Ramachandran plots (Figure 5.3 and Figure 5.S3).
Figure 5.3 Temperature dependence of properties of a single ELP chain: (a) Interaction energy of each residue in an ELP pentamer with water. (b) Number of hydrogen bonds formed between each residue in an ELP pentamer and water. (c) SASA of each residue in an ELP pentamer.
Figure 5.4 (a) Probability distribution of the size $N_{\text{max}}$ of the largest water network around polypeptide. (b) A snapshot of (VGPVG)$_{18}$ and the largest water network around it at 350 K.
Figure 5.5 Interaction between two ELPs at 350 K. (a) Distance between center-of-mass of two polypeptides and (b) peptide-peptide interaction energy (c) Number of inter-chain hydrogen bonds.
Figure 5.S1 Temporal evolution of radius of gyration (Rg) of (VPGVG)18 at different temperatures. Representative structures of (VPGVG)18 from cluster analysis of MD trajectories at (a-c) 290 K and (d-f) 350 K are displayed.
Figure 5.S2 Temporal evolution of radius of gyration ($R_g$) of (VGPVG)$_{18}$ at different temperatures. Representative structures of (VGPVG)$_{18}$ from cluster analysis of MD trajectories at (a-c) 290 K and (d-f) 350 K are displayed.
Figure 5.S3 Time autocorrelation function of (VPGVG)$_{18}$ backbone C$\alpha$- C$\alpha$ bond vectors: (a) unit vector along backbone C$\alpha$- C$\alpha$ bond; (b) unit vector out of plane. We can estimate the relaxation time of (VPGVG)$_{18}$ backbone vectors from the time autocorrelation profiles which is less than 25 ns for each group.
Figure 5.S4 Time autocorrelation function of (VGPVG)$_{18}$ peptide backbone C$\alpha$- C$\alpha$ bond vectors: (a) unit vector along backbone C$\alpha$- C$\alpha$ bond; (b) unit vector out of plane. We can estimate the relaxation time of (VPGVG)$_{18}$ backbone vectors from the time autocorrelation profiles which is less than 15 ns for each group.
Figure 5.S5 (a) Temperature dependence of Number of the intramolecular intra-peptide hydrogen bonds ($n_{pp}$) and intramolecular water-mediated peptide hydrogen bonds ($n_{pwp}$) of (VPGVG)$_{18}$ and (VGPVG)$_{18}$. (b) Temperature dependence of the number of water molecules in the first watershell of (VPGVG)$_{18}$ and (VGPVG)$_{18}$. 
Figure 5.S6 Ramachandran plot (\(\phi, \psi\) distributions) for each residue in a pentamer, (a) VPGVG at 290 K, (b) VPGVG at 350 K, (c) VGPVG at 290 K, (d) VGPVG at 350 K, colored by intensities. The most populated regions on the Ramachandran plots are labelled.
Figure 5.S7 Average frequency of occurrence of each residues in (VPGVG)$_{18}$ (a, b) and (VGPVG)$_{18}$ (c, d) in each type of secondary structural motifs at 290 K (a, c) and 350 K (b, d).
Chapter 6 Million-atom Molecular Dynamics Simulation on the Aggregation Behaviors of Elastin-like Polypeptides

6.1 Introduction

There has been great interest in the LCST behaviors of elastin-like polypeptides (ELP) in aqueous environment\(^1,2\). In Chapter 5, we studied the role of amino acid order on the structural and dynamical changes of the two model ELP chains, poly(VPGVG) and poly(VGPVG), and their hydration properties. Through the examination of the behaviors of single polypeptides in a range of temperatures in aqueous solution, we found that the switch of Pro-Gly into Gly-Pro still leads to a functional polymer which exhibited LCST behaviors, but destroy the β-turns which resulting a more extended conformation and a larger solvent exposure. Herein, we conducted atomistic MD simulations to show the evolution of structural properties of ELP molecules during the aggregation process of multiple polypeptides at elevated temperature. To the best of our knowledge, no attempt has yet become successful in the prediction of the structural dynamics of the ELP aggregation directly from MD simulations. Such a direct atomistic MD simulation should be able to provide valuable insights into various properties of ELP molecules within an aggregate, and it may also unveil the nature of the LCST behaviors of ELPs beyond the single molecular description.
6.2 Materials and Methods

Fully atomistic molecular dynamics simulations were performed using Amber 12.0\textsuperscript{3} and the ff99SB force field for proteins with explicit solvent using the TIP3P water model.\textsuperscript{4} All simulations were performed using an NPT ensemble with periodic boundary conditions applied at 350 K. The lowest-energy representative single polypeptide structures obtained in the previous study reported in Chapter 5 were then taken to be the seed structures in the simulations of the aggregation of 27 polypeptides, which were uniformly distributed in the simulation box (Figure 1). These polypeptides were then solvated by TIP3P water molecules. The simulation box contains 360,774 water molecules with the size of 22.5 nm × 23.4 nm × 22.9 nm for multi-poly(VPGVG) system and 354.491 water molecules with the size of 22.8 nm × 23.4 nm × 22.3 nm for multi-poly(VGPVG) system. Equilibration protocols and MD simulations were the same as described for single-peptide simulations\textsuperscript{5}. The production simulations were carried out for more than 700 ns with a 2 fs time step. MD simulations were carried out on graphic processing unit (GPU), which enabled handling of ~1 million atoms in both systems.

6.3 Results and Discussions

It was proposed that the ELPs form coacervate droplets above the LCST due to the collapse and aggregation of polypeptide molecules.\textsuperscript{6} Herein, we conducted MD simulations on the aggregation process of 27 poly(VPGVG) and 27 poly(VGPVG) in explicit water. We observed that multiple polypeptides which were arranged in water at 350 K aggregated and
formed one cluster after 300 ns simulation, as shown in Figure 6.1, which can be used to mimic the properties of an organic-rich phase of the ELP systems. Accordingly, various properties of the polypeptide clusters can be studied. The averaged radius of gyration of each polypeptide as well as the averaged intra-chain hydrogen bonds were calculated (Figure 6.2). We find that during the aggregation process, both poly(VPGVG) and poly(VGPVG) chains extend due to the inter-chain interplay.

The number of intermolecular hydrogen bonds between ELPs has been calculated during the simulation, and the results are presented in Figure 6.3. In our simulations, we found that there was an increasing number of hydrogen bonding interactions between polypeptides. This observation indicates that interchain hydrogen bonding stabilizes the aggregates and gives the most significant contribution to the formation of ELP aggregates.

We also examined the Ramachandran plots for each residue in the repeated pentamer VPGVG and VGPVG in the aggregates taken at 0 ns and 700 ns of the MD simulation and estimated the propensities of each secondary structural motifs. The high intensity region in Ramachandran plot represents the most favorable, low-energy torsion angles for each residue. Herein, we compared the Ramachandran plots taken at the beginning and the end of simulations and observed similarities but also significant differences. As showed in Figure 6.4, we observed the changes in the torsion angle of Gly3 in poly(VPGVG) over aggregation process. One high intensity region labeled by a_1 in VPGVG is missing in the plot for Gly3 in the aggregate, indicating the omittance of Type 1 β-turn structures. Distinct changes were also observed for Val1 and Val4 for poly(VGPVG). Although Val1 in VGPVG occupy
similar conformational spaces before and after aggregation in Ramachandran plots, the most populated angles adopted by Val1 in VGPVG changing from $b_1 (-127.4^\circ, -7.5^\circ)$ to $c_1 (-135.8^\circ, 147.6^\circ)$, indicating the decreasing trend in $\beta$-strands. Also the region labelled by $b_2$ for Val4 in VGPVG is missing in the plot for Val4 in the aggregate, indicating the decrease in helical structures after aggregation.

Tables 6.1 and 6.2 show the averaged propensities of secondary structural motifs for polypeptides before and after aggregation, which were determined by a secondary structure recognition algorithm DSSP method based on H-bonding patterns\textsuperscript{7}. The secondary structural differences as a result of the ELP aggregation were revealed by million-atom MD simulations. During the aggregation, poly(VPGVG)s lose Turn and $\beta$-strand structures; whereas for poly(VGPVG)s we observed an increase in $\beta$-strand and decrease in helical secondary structure motifs, which are also consistent with the observations from Ramachandran plots.

In summary, we conducted million-atom molecular dynamics simulations to study the aggregation properties of 27 ELPs at a fully atomistic level. We found that during the simulation all 27 polypeptides aggregated into one cluster. The formation of the peptide aggregates causes essential increase in the radius of gyration of the peptides, which may be attributed to the decrease in the intra-chain hydrogen bonding and the increase in the inter-chain hydrogen bonding. We also observed the changes in torsion angles and the propensities of secondary structural motifs over the aggregation process in both systems. The atomistic structure of ELP aggregates reported in this work will be useful for elucidating the
mechanism for the spontaneous organization of such ELPs behind the LCST phase transition processes.

Funding Sources

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107).

ACKNOWLEDGMENT

This work was supported by the NSF's Research Triangle MRSEC (DMR-1121107). The computer support was provided by the High Performance Computing Center at North Carolina State University.
6.4 References


Table 6.1 Secondary structure propensities for (VPGVG)18 at 0 ns and 700 ns

<table>
<thead>
<tr>
<th></th>
<th>0 ns</th>
<th>700 ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>0.21±0.04</td>
<td>0.17±0.05</td>
</tr>
<tr>
<td>β-strand</td>
<td>0.14±0.03</td>
<td>0.11±0.04</td>
</tr>
<tr>
<td>Helix</td>
<td>0.00758±0.01491</td>
<td>0.00611±0.01306</td>
</tr>
</tbody>
</table>

Table 6.2 Secondary structure propensities for (VGPVG)18 at 0 ns and 700 ns

<table>
<thead>
<tr>
<th></th>
<th>0 ns</th>
<th>700 ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn</td>
<td>0.14±0.05</td>
<td>0.12±0.06</td>
</tr>
<tr>
<td>β-strand</td>
<td>0.06±0.03</td>
<td>0.10±0.05</td>
</tr>
<tr>
<td>Helix</td>
<td>0.053±0.033</td>
<td>0.041±0.031</td>
</tr>
</tbody>
</table>
Figure 6.1 Snapshots of 27 (VPGVG)_{18} at (a) 0 ns and (b) 700 ns, and the snapshot of 27 (VGPVG)_{18} at (c) 700 ns. (d) Number of polymer clusters in the simulation system as a function of simulation time.
Figure 6.2 Temporal evolution of (a) the averaged radius of gyration of polypeptides and (b) the averaged number of intra-chain hydrogen bonds of polypeptides.
Figure 6.3 Number of inter-chain hydrogen bonds as a function of simulation time.
Figure 6.4 Ramachandran plot ($\phi$, $\psi$ distributions) for each residue in repeated pentamer in (a) (VPGVG)$_{18}$ at 0 ns, (b) (VPGVG)$_{18}$ at 700 ns, (c) (VGPVG)$_{18}$ at 0 ns, and (d) (VGPVG)$_{18}$ at 700 ns, colored by intensities.
Chapter 7 An implicit solvent ionic strength (ISIS) method to model polyelectrolyte systems with dissipative particle dynamics

Nan K. Li, William H. Fuss, Yaroslava G. Yingling*

Published in Macromolecular Theory and Simulations 24 (1), 7-12

Herein, a new coarse-grained methodology for modeling and simulations of polyelectrolyte systems using implicit solvent ionic strength (ISIS) with dissipative particle dynamics (DPD) is presented. This ISIS model is based on mean-field theory approximation and the soft repulsive potential is used to reproduce the effect of solvent ionic strength. The capability of the ISIS model is assessed via two test cases: dynamics of a single long polyelectrolyte chain and the self-assembly of polyelectrolyte diblock copolymers in aqueous solutions with variable ionic strength. The results are in good agreement with previous experimental observations and theoretical predictions, which indicates that our polyelectrolyte model can be used to effectively and efficiently capture salt-dependent conformational features of large-scale polyelectrolyte systems in aqueous solutions, especially at the salt-dominated regime.

7.1 Introduction

Polyelectrolytes are macromolecules with a large number of ionizable groups, which can dissociate into charged polymer chains and small counterions in polar solutions. The
strength of the electrostatic interactions may be moderated by increasing the concentrations of added salt due to a screening effect, which results in changes in the dimensions and dynamics of the polyelectrolyte chain.\textsuperscript{1} Polyelectrolyte block copolymers, formed with synthetic or natural (e.g., DNA, RNA and peptide) ionic blocks linked to a hydrophobic block, have attracted considerable attention due to widespread industrial and biomedical applications\textsuperscript{2}. However, modeling of complex polyelectrolyte systems can be challenging and computationally intensive. Atomistic or coarse-grained models where the solvent and ions are considered explicitly are limited by their high computational cost, especially as the system size increases, due to the greater number of atoms/atom groups introduced.

Dissipative Particle Dynamics (DPD) is a particle-based meso-scale simulation technique, which has been used as a tool to simulate mesoscale phenomena in a wide range of applications such as dilute solution polymer behavior\textsuperscript{3}, phase separation\textsuperscript{4}, polymer brush coatings\textsuperscript{5}, and self-assembly from solution\textsuperscript{6}. DPD technique has been confirmed to be capable of capturing essential features of macromolecules in various environments.\textsuperscript{7} The incorporation of electrostatic interactions into DPD, which was originally proposed by Groot\textsuperscript{8}, introduced a fascinating tool for the study of polyelectrolytes.\textsuperscript{7a,b,9} Groot’s model is based upon the particle-particle particle-mesh (PPPM) method with a charge distribution function adapted to the use of a soft potential. The second approach for the calculation of long range electrostatic interactions was developed by Gonzáles-Melchor and co-workers\textsuperscript{10}, using the Ewald technique\textsuperscript{11} with a charge distribution on the DPD beads. DPD is extremely fast in comparison with molecular dynamics (MD), but the implementation of long range
electrostatic interactions considerably increases the computational cost and complexity of DPD. Thus, there is a need to develop a new type of methodology with incorporated electrostatic interactions to achieve a good balance between reasonable physical description and computational feasibility. Herein, we present a new coarse-grained methodology for the modeling and simulation of polyelectrolyte systems with Implicit representation of Solvent Ionic Strength (ISIS). A DPD model with the soft repulsive potential to reproduce the effect of solvent ionic strength is utilized to explore the properties of single polyelectrolyte chain and the self-assembly of polyelectrolyte block copolymers in aqueous solution.

7.2 Method

In DPD, a number of molecular entities are coarse-grained into an element, thereafter called a DPD bead. These DPD beads are subject to soft potentials and governed by predefined collision rules. The beads move according to Newton’s equations of motion. Thus, for a DPD bead \( i \), we have

\[
\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i , \quad m_i \frac{d\mathbf{v}_i}{dt} = \sum_{j \neq i} f_{ij} , \quad f_{ij} = (F_{ij}^C + F_{ij}^D + F_{ij}^R) + f_{ij}^S \tag{7.1}
\]

where \( m_i, \mathbf{r}_i \) and \( \mathbf{v}_i \) are the mass, position, and velocity of bead \( i \), respectively. \( f_{ij} \) is the interbead force on bead \( i \) by a bead \( j \). \( F_{ij}^C, F_{ij}^D, \) and \( F_{ij}^R \) are the conservative force, dissipative force, and the random force, respectively, acting between beads \( i \) and \( j \). They are given by
where $a_{ij}$ is the maximum repulsion between beads $i$ and $j$; $r_{ij} = r_i - r_j$, $r_{ij} = |r_{ij}|$, $v_{ij} = v_i - v_j$, and $\bar{r}_{ij} = r_{ij}/r_{ij}$ is the unit vector directed along $j$ to $i$. Coefficients $\gamma$ and $\sigma_D$ characterize the strengths of the dissipative and random forces, where $\gamma = \frac{\sigma_D^2}{2k_B T}$. $\Delta t$ is the iteration time step, and $\theta_{ij}(t)$ is a symmetric random variable. The two weight functions and their prefactors must satisfy the following relationship:\footnote{13}:

$$w^D(r_{ij}) = [w^R(r_{ij})]^2 = \begin{cases} \left(1 - \frac{r_{ij}}{r_c}\right)^2, & r_{ij} < r_c, \\ 0, & r_{ij} \geq r_c \end{cases}$$

(7.3)

In our system, polymer chains are represented with a bead-spring type particle model, where adjacent beads in the chains are connected via an extra harmonic spring: $F_{ij}^S = C r_{ij}$, where $C$ is the spring constant. The value for intrachain spring constant is set to 4.0 between hydrophobic beads and 50.0 between polyelectrolyte beads. The choice of $C$ will not affect the qualitative behavior of the system. Water is also modeled as coarse-grained beads, each bead corresponding to a group of several $\text{H}_2\text{O}$ molecules. Also in DPD, the properties of the system are expressed using dimensionless quantities in units of the cutoff $r_c$, the energy scale $k_B T$ and the bead mass $m_0$. As a consequence, the unit of time, $\tau$, is $\tau = \sqrt{r_c^2 m_0 / k_B T}$. 

182
The repulsion parameters in our system were parameterized based on the Groot and Warren\textsuperscript{12b} scheme, where the repulsion parameter for DPD beads of water is equal to $a_{ii} = 25$ and the bead density of the system, $\rho = 3$, was chosen to correspond to a liquid at room temperature. The conservative force determines the thermodynamics of the DPD system, and is defined by a purely repulsive (parabolic) soft-core potential. The DPD conservative forces produces an equation of state (EOS) that can be approximated by the equation $p = \rho k_B T + \alpha a \rho^2$ ($\alpha = 0.101 \pm 0.001$), as the virial expansion with pressure $p$, density $\rho$, and repulsive parameter $\alpha = a_{ij}$.\textsuperscript{12b} Note that such an EOS is quite different from the EOS of a real fluid, so the DPD EOS breaks down for $\rho \to 0$.\textsuperscript{14} This approximation, however, is very good for $\rho \geq 3$ and $a \geq 15$.\textsuperscript{12b} A good approximation for the soft potential parameter $\alpha$ that holds for sufficiently high density could be $\alpha \sim \nu$, where $\nu$ is the second virial coefficient.

The standard soft potential is obtained by matching with experimental or atomistic results\textsuperscript{12b}

$$\kappa^{-1} = \frac{1}{k_B T} \frac{\partial p}{\partial \rho} = \frac{1}{k_B T} \frac{\partial p}{\partial n} \frac{\partial n}{\partial \rho} \quad (7.4)$$

By taking the dimensionless compressibility of water ($\kappa^{-1} \sim 16$) into account, the repulsion parameter becomes:

$$a_{ii} = 75 \frac{k_B T}{\rho r_c^2} \quad (7.5)$$

Thus, if $\rho = 3$ the repulsion parameter between beads of the same type is $a_{ii} = 25$. 

183
In the case of polyelectrolytes in a salt-dominated environment, i.e., where the bulk concentration of added salt exceeds the concentration of counterions, the short-range repulsion between charged monomers is governed by the screened (by salt) Coulomb interactions and complemented by the non-electrostatic (excluded volume) interactions. Based on the mean-field approximation, the effective second virial coefficient calculated per monomer is then equal to\(^1\):

\[
v = v_A + \alpha^2/c_s\]

(7.6)

where \(v_A \leq 1\) is the bare non-electrostatic contribution to the second virial coefficient, \(c_s\) is the solvent ionic strength, and \(\alpha\) is the degree of ionization.

Analogous to the second virial coefficient, \(v\), we represent the repulsive parameters between polyelectrolyte beads as:

\[
a_{pp} = a_{ii} + a_{e\text{lec}}\]

(7.7)

where \(a_{e\text{lec}} \sim c_s^{-1}\) and \(a_{ii} = 25\) (at \(\rho = 3\)). Thus, an increase in the solvent ionic strength \(c_s\) leads to \(a_{pp} \rightarrow a_{ii}\), which reduces the behavior of polyelectrolytes to that of neutral polymers with the same solvation properties. As a result, \(a_{pp}\) in our model is inversely proportional to the solvent ionic strength. Similar to Ref.\(^{15a}\), our model assumes that the polyelectrolyte assemblies have the same degree of ionization as a single chain in solution.

Additionally, increasing the repulsive parameters between polyelectrolyte beads can reduce the bond crossing effect, which is beneficial for the simulation of dense system. However, as for the system with high solvent ionic strength, bond-crossing cannot be
avoided, especially when $a_{ii} = 25$, where salt ions can screen all the charges on the polyelectrolyte chain. To avoid this effect, an extra “spring-spring repulsion” can be applied.

In order to test the effectiveness of our model, we considered two test cases: (1) dynamics of a single polyelectrolyte in aqueous solution and (2) self-assembly of diblock copolymers, where each diblock copolymer is composed of a polyelectrolyte and a hydrophobic block, in aqueous solution as a function of solvent ionic strength. All simulations were carried out with the LAMMPS program package (Sandia National Laboratory).\(^\text{16}\)

In the first case, a single polyelectrolyte chain composed of 300 beads was placed in a cubic box ($40 \times 40 \times 40 r_c^3$) containing 192,000 beads. To represent the change in salt concentration, we varied the repulsive parameters polyelectrolyte beads, $a_{pp}$, from 25 to 60, where $a_{pp} = a_{ii} + a_{elec}$. Specifically, $a_{elec}$ was varied from 0 to 35 with $a_{ii}$ fixed at 25. To examine the conformational properties of a single polyelectrolyte chain at different solvent ionic strengths, several parameters, namely radius of gyration, $R_g$, end-to-end distance, $D_{ee}$, and persistence length, $l_p$, were calculated and compared to experimental observations and theories.

In the second case, 300 diblock chains composed of a longer hydrophilic polyelectrolyte block ($N_A=30$) and a shorter hydrophobic block ($N_B=4$) were placed in a cubic box $36 \times 36 \times 36 r_c^3$ (Figure 7.2a). Similar to the first case, we varied the repulsive parameter between polyelectrolyte beads, $a_{pp}$, from 25 to 60. To obtain aggregates in water, the repulsive parameters for the polyelectrolyte-hydrophobic segment interactions and water-
hydrophobic interactions were set to 90 and 100, respectively. The interaction parameter between the hydrophilic polyelectrolyte segments and water molecules was set to 26. DPD trajectories were processed using in-house scripts to calculate the aggregation number, $P$, and anisotropy, $\kappa^2$, with all the aggregates identified based on a general distance criterion (see Supporting Information).

7.3 Results and Discussion

7.3.1 Single polyelectrolyte

Within the approximation of uniform swelling of the polyelectrolyte chains, the repulsive parameter between polyelectrolyte beads, $a_{pp}$, accounts for the excluded volume and electrostatic interaction between polyelectrolyte beads, as the screened electrostatic becomes short-ranged at high ionic strength solvent. The flexibility of a polymer in solution can be characterized by the end-to-end distance, $D_{ee}$, which is plotted against the repulsive parameter $a_{pp}$ in Figure 7.1a. We observed that the $D_{ee}$ decreased as $a_{pp}$ decreased. This behavior is expected, since adding salt ions to polyelectrolyte solutions aids in the screening of the repulsion between the charged segments and consequently reduces the end-to-end distance. With a higher repulsive parameter (low salt concentration), a plateau in the $D_{ee}$ profile is observed, which reflects the competition between the connectivity of polymer chains and the electrostatic repulsion between charged monomers. To obtain scaling relationship, a log-log plot of $D_{ee}$ vs. $a_{elec}$ is shown in Figure 7.1b. For a single long polyelectrolyte, a scaling relation of $D_{ee} \sim a_{elec}^{0.20}$, i.e., $D_{ee} \sim c_s^{-0.20}$ was obtained. The
simulation results show that $R_g$ (Figure 7.1c) also follows a scaling relationship with the solvent ionic strength $R_g \sim c_s^{-0.18}$. The calculated exponent is comparable to the prediction from a recent MD simulation of polyelectrolytes in which a scaling dependence $R_g \propto c_s^{-0.2}$ for polyelectrolyte in dilute solution and $R_g \propto c_s^{-0.125}$ at high ionic strength solvent in semi-dilute solution were presented\textsuperscript{17}.

The dependence of polyelectrolyte persistence length, $l_p$, on solvent ionic strength is related to the repulsive electrostatic interactions, the screening effects, and their mutual effects on the thermodynamics properties of these systems. It is known that electrostatic repulsion between the ionized groups stiffens the polyelectrolyte chain, which results in a larger value of persistence length than that of a neutral chain. At relatively high solvent ionic strength, the screening effects brought by the excess of the counterions become more important. The dependence of polyelectrolyte persistence length on the solvent ionic strength is still being discussed. For example, the Odijk-Skolnik-Fixman (OSF) theory shows that the persistence length scales with the inverse of the solvent ionic strength\textsuperscript{18}, $l_p^{\text{OSF}} \propto c_s^{-1}$. The same scaling relation was observed in several computer simulation studies\textsuperscript{19}. However, other theoretical studies have shown a slightly different relation of the persistence length of a flexible polyelectrolyte, $l_p \propto c_s^{-0.5}$\textsuperscript{20}. The same relation was obtained by several experimental studies\textsuperscript{21} and MD simulations\textsuperscript{17}. It was proposed that OSF relationship holds well for low solvent ionic strengths and for relatively stiff polyelectrolytes. As for highly flexible chains at high solvent ionic strengths, a weaker dependence can be observed.\textsuperscript{21a}
In our ISIS DPD model a scaling relation of \( l_p \propto c_s^{-0.375} \) was obtained (Figure 7.1d). When \( a_{pp} = a_{ii} = 25 \) the solvent ionic strength is so high that \( l_p \) is reduced to that of a neutral chain. The weaker dependence of persistence length on solvent ionic strength in our model as compared to previous studies is probably due to the underestimation of the excluded volume contribution at relatively low solvent ionic strength and the high flexibility of a polymer chain in a DPD system. Overall, the observed dynamics of a single polyelectrolyte chain as a function of salt concentration in our model shows a general agreement with the developed theories and experimental observations.

### 7.3.2 Self-assembly of polyelectrolyte block copolymers

Next, we examined a more complicated case where the polyelectrolyte was a part of a diblock copolymer with a hydrophobic end. Figure 7.2 shows the dynamics of self-assembly of diblock copolymer chains at \( a_{pp} = 40 \) and a set of final snapshots of the self-assembled micelles as a function of repulsive parameters between polyelectrolyte beads, \( a_{pp} \), (or inverse salt concentration). By varying solvent ionic strength the formation of three different kinds of micellar morphologies was observed, namely, spherical, cylindrical and wormlike. As the repulsive parameter \( a_{pp} \) decreases from 60 to 25 (i.e., increase in salt concentration), the transition from spherical micelles to cylindrical micelles to wormlike aggregates was observed. This observation is consistent with experimental studies and theoretical predictions, indicating that an increase in solvent ionic strength gives rise to a the spherical-to-cylindrical/wormlike transition for the micelles of polyelectrolyte diblock copolymers.
The micellar morphology can be examined in terms of the anisotropy, $\kappa^2$, defined in Supporting Information. Figure 7.3a shows the probability distribution of anisotropy $\kappa^2$ of micellar cores. The $a_{pp}=25$ case (high salt concentration) was excluded from this analysis because only a single worm-like micelle was formed. The profile for $a_{pp} = 30$ not only has a higher mean but also is much wider than the other profiles indicating the formation of polydisperse cylindrical micelles at high solvent ionic strength. At lower solvent ionic strength ($a_{pp} > 30$), most of the micelles obtained are spherical.

The polydispersity of the micelles was assessed based on the distribution of the aggregation number, $P$, as shown in Figure 7.3b. The aggregation number decreased as the strength of repulsive interactions $a_{pp}$ increased (due to reduced salt ions). When $a_{pp} = 30$, a wide distribution of micelles was obtained with aggregation numbers ranging from $P = 2$ to 74. As $a_{pp}$ increased, the distribution of $P$ became narrower, which is consistent with theoretical predictions showing that the most probable aggregation number for spherical micelles is close to the average aggregation number.\textsuperscript{15b} Overall, the results of our simulations of the self-assembly of diblock polyelectrolyte copolymers into micelles are consistent with experimental observations and theoretical predictions.

### 7.4 Conclusion

Herein, we present a new simple coarse-grained methodology for modeling polyelectrolyte systems with dissipative particle dynamics. Our ISIS model approximates the repulsive parameter between polyelectrolytes in DPD with a second virial coefficient
formalism in good solvent. We assessed our model with two test cases, namely, a single long polyelectrolyte chain and the self-assembly of polyelectrolyte diblock copolymers in aqueous solutions.

The main advantages of our model are its computational speed and simplicity of implementation. The simulation results suggest that this DPD simulation scheme gives a route wherein one can effectively and efficiently capture the salt dependence of conformational features of polyelectrolytes in aqueous solutions, especially at the salt-dominated regime, (i.e., bulk concentration of added salt exceeds the concentration of counterions trapped inside the polyelectrolyte region and neutralizing the bare charge of the chain). Also, this simulation scheme enables the intensive study of the self-assembly behaviors of polyelectrolyte block copolymers at high solvent ionic strength. Such a development in the coarse-grained representation of polyelectrolytes opens avenues to study complex systems containing polyelectrolytes in an efficient way. Moreover, our test cases show an overall agreement with the previous experimental observations and theoretical predictions.

The main limitation of our model is that the radial gradients in the polyelectrolyte density and mobile ion distribution are neglected due to the mean-field approximation and local electroneutrality approximation made in the derivation of this methodology. However, this effect is negligible when the ionization constant of the polyelectrolyte chain and/or solvent ionic strength are high. This model should be used with caution for weak polyelectrolytes at low solvent ionic strengths. Overall, this model best suited for studies of
processes and properties of various systems containing strong polyelectrolytes (DNA, RNA, poly(styrenesulfonate), etc.) in high solvent ionic strength.

Supporting Information
Supporting Information is available from the Wiley Online Library

Acknowledgements
This work was supported by the NSF’s Research Triangle MRSEC (DMR-1121107) and NSF CMMI-1150682. The computer support was provided by the High Performance Computing Center at North Carolina State University.
7.5 References


192


Figure 7.1 Properties of a single long polyelectrolyte chain as a function of solvent ionic strength. (a) End-to-end distance as a function of the repulsive parameter, $a_{pp}$; (b) Logarithmic plot of the end-to-end distance as a function of $a_{elec}$; (c) Radius of gyration as a function of $a_{elec}$; and (d) Dependence of the chain persistence length, $l_p$, on the electrostatic repulsive parameter, $a_{elec}$. Note that $a_{elec} \sim c_s^{-1}$. 
Figure 7.2 Snapshots of micellization of polyelectrolyte copolymers in DPD simulations. (a-d) Timeline of a 300-chain system with $a_{pp} = 40$. (a) Time step = 1, (b) Time step = 1.0E5, (c) Time step = 1.0E6, and (d) Time step = 4.0E7. (e-h) Typical snapshots of polyelectrolyte block copolymers in aqueous solution at (e) $a_{pp} = 25$, (f) $a_{pp} = 30$, (g) $a_{pp} = 50$, and (h) $a_{pp} = 60$. The yellow and blue beads represent the hydrophobic block and hydrophilic block, respectively. Water is omitted for clarity.
Figure 7.3 The properties of polyelectrolyte diblock micelles as a function of solvent ionic strength. (a) Probability distribution of the anisotropy of the micellar core and (b) Probability distribution of the aggregation number.
Chapter 8 Prediction of solvent-induced morphological changes of polyelectrolyte diblock copolymer micelles

Nan K. Li, William H. Fuss, Lei Tang, Renpeng Gu, Ashutosh Chilkoti, Stefan Zauscher, Yaroslava G. Yingling

Published in Soft matter 11 (42), 8236-8245

Self-assembly processes of polyelectrolyte block copolymers are ubiquitous in industrial and biological processes; understanding their physical properties can also provide insights into the design of polyelectrolyte materials with novel and tailored properties. Here, we report systematic analysis on how the ionic strength of the solvent and the length of the polyelectrolyte block affect the self-assembly and morphology of the polyelectrolyte block copolymer materials by constructing a salt-dependent morphological phase diagram using an implicit solvent ionic strength (ISIS) method for dissipative particle dynamics (DPD) simulations. This diagram permits the determination of the conditions for the morphological transition into a specific shape, namely vesicles or lamellar aggregates, wormlike/cylindrical micelles, and spherical micelles. The scaling g behavior for the size of spherical micelles is predicted, in terms of radius of gyration ($R_{g,m}$) and thickness of corona ($H_{corona}$), as a function of solvent ionic strength ($c_s$) and polyelectrolyte length ($N_A$), which are $R_{g,m} \sim c_s^{-0.06} N_A^{0.54}$ and $H_{corona} \sim c_s^{-0.11} N_A^{0.75}$. The simulation results were corroborated through AFM and static light scattering measurements on the example of the self-assembly
of monodisperse, single-stranded DNA block-copolynucleotides (polyT50-b-F-dUTP). Overall, we were able to predict the salt-responsive morphology of polyelectrolyte materials in aqueous solution and show that a spherical-cylindrical-lamellar change in morphology can be obtained through an increase in solvent ionic strength or a decrease of polyelectrolyte length.

8.1 Introduction

Polyelectrolyte diblock copolymers (PDCs), which combine the properties of polyelectrolytes (i.e., sensitive to changes in solvent ionic strength and pH) with those of surfactants, can self-assemble in an aqueous environment into a variety of responsive morphologies, including spherical micelles, star-like/hairy and crew-cut, cylindrical micelles, vesicles, lamellar mesophases, and micellar aggregates. Such assemblies are promising candidates as carriers for drug and gene delivery, where the morphology and the size of the assemblies determine their transport properties and delivery capabilities. For example, it was suggested that cylindrical delivery vehicles display a longer circulation time in the body than spherical carriers, due to reduced interactions with the blood vessel walls. Aside from biomedical applications, preparation of micelles with various morphologies is also important in the fabrication of soft lithography templates. Ability to predict and control the responsive morphology of PDCs is needed for the development of new materials systems with a broad range of biological and technological applications.
Many experimental studies have been focused on the investigation of the structure and properties of micellar solutions of various polyelectrolyte block copolymers.\textsuperscript{6} It was determined that increasing the solvent ionic strength induces the shrinkage of corona chains, increases the micellar aggregation number, and alters the micellar morphology.\textsuperscript{1c, 7} However, the mechanism controlling self-assembly, aggregation and overall morphology of polyelectrolyte block copolymers is complex and is largely driven by the characteristically intricate equilibrium of non-covalent interactions, including electrostatic, steric, hydrophobic, Van der Waals interactions, and hydrogen bonding.\textsuperscript{8} With the rapidly growing ability to synthesize PDCs with specific size and dispersity, it is important to develop a comprehensive method that is able to predict the solvent-dependent properties, morphologies, and assembly kinetics of these molecular assemblies.

Mean-field theories and scaling theories can be used to describe the properties of PDC micelles.\textsuperscript{9} One of the most notable theories is the one developed by Borisov and Zhulina which is based on the mean-field theory.\textsuperscript{8, 9c} In this theory, the free energy of the micelles is described in terms of the degree of polymerization of the polyelectrolyte and hydrophobic blocks, and several external parameters that control the interaction strength, such as the second virial coefficient, which represents non-electrostatic excluded-volume interactions, and the Coulomb interactions between the charged monomers. Although, morphological transitions were predicted through the free energy expressions for spherical, cylindrical, or lamellar micelles, according to their respective packing geometries.\textsuperscript{8, 9c}
theoretical analysis of the self-assembly and morphology of PDCs is challenging compared to that of neutral systems and more developments in this area are needed.\textsuperscript{10}

Computational techniques can be used to elucidate the processes and mechanisms of the self-assembly of amphiphilic components in solution and are a valuable complement to experimental and theoretical approaches.\textsuperscript{11} However, the use of computational and modeling techniques for self-assembly of polyelectrolyte block copolymers in aqueous solution has not been extensively explored, despite the significant interest in tailoring polyelectrolytes systems for various applications. All-atom simulations of the self-assembly of many polyelectrolyte chains are computationally expensive, because the size of the system, as well as the equilibration and residence times of counterions, can be prohibitively large.\textsuperscript{12} Therefore, to achieve a balance between reasonable physical description and computational feasibility, a commonly used strategy to simulate these processes is to use coarse-grained modeling techniques, such as coarse-grained molecular dynamics (MD) or Dissipative Particle Dynamics (DPD).\textsuperscript{13} However, modeling of complex polyelectrolyte systems in aqueous solution is challenging and computationally intensive even for coarse-grained models due to the implementation of long range electrostatic interactions.\textsuperscript{14} Only a few of such simulation studies have been reported where the system size was relatively small and the interactions between a scarce number of polyelectrolyte molecules were investigated.\textsuperscript{15} In terms of predicting and observing morphological changes of PDCs as a function of ionic strength and pH, the most notable study is the recent work by Pantano \textit{et al.} where coarse-grained MD simulations were used to reproduce different morphologies of short, charged
diblock copolymers due to changes in pH and the presence of Ca\textsuperscript{2+}.\textsuperscript{16} To the best of our knowledge there are no comprehensive, large-scale comprehensive computational studies on the prediction of self-assembly and morphology of PDC micelles with polyelectrolyte corona.

To overcome the current computational limitations, we have recently developed a new implicit solvent ionic strength (ISIS) methodology for the DPD method which combines the explicit solvent with the implicit ionic strength representation and permits large-scale simulation of self-assembling polyelectrolytes and their response to the changes in ionic strength, especially at the salt-dominated regime.\textsuperscript{17} Briefly, in our model we parameterize the repulsive non-bonded parameter ($a_{pp}$), between and within the polyelectrolyte chains in analogy with a second virial coefficient formalism for a good solvent based on mean-field theory for polyelectrolyte systems. Our coarse-grained modeling methodology is designed for studies of large-scale systems that involve the ionic strength-dependent behavior of strong polyelectrolytes (\textit{i.e.} DNA, RNA, poly(styrenesulfonate), \textit{etc.}), which is currently unreachable by other computational methods, and enables us to obtain important descriptors of micelle morphology and structure, \textit{i.e.}, the aggregation number, the corona and core sizes, and micelle anisotropy, as a function of polyelectrolyte block length and solvent ionic strength.

In this paper, we use our ISIS DPD model to predict the ionic strength-dependent morphology of highly asymmetric, PDCs in aqueous solution. Specifically, we explore the role of polyelectrolyte block length and solvent ionic strength on the morphological
properties of the assemblies. The diblocks have a short hydrophobic block with a fixed
degree of polymerization, $N_B = 4$, and a variable length of the hydrophilic block with a degree
of polymerization, $N_A$, ranging from 4 to 90 (Figure 8.1a). To validate our predictions, we
examine the self-assembly of asymmetric, 54-mer single stranded DNA (ssDNA)
amphiphiles using static light scattering (SLS) and AFM imaging, and compared the results
to the simulation predictions. These polyelectrolyte amphiphiles were synthesized by
appending on average about four hydrophobic nucleotides (*i.e.*, Fluorescein -dUTP) to the 3’
termini of hydrophilic polynucleotides (polyT$_{50}$) via an enzymatic polymerization reaction.$^{11a}$
In aqueous solution, the hydrophobic F-dUTP segments aggregate and form the micellar
core. Overall, we demonstrated the ability to predict the morphology of PDCs in aqueous
solution and their response to a change in solvent ionic strength.

### 8.2 Experimental Section

#### 8.2.1 Simulation Method

All simulations were performed using Dissipative Particle Dynamics (DPD) via
LAMMPS$^{18}$. DPD is a coarse-grained simulation technique in which one DPD bead
represents a group of atoms or a volume of fluid that is large on the atomistic scale but still
macroscopically small. $^{19}$ All beads move according to Newton’s equations of motion,
$$m_i \frac{dv_i}{dt} = \sum_{j \neq i} f_{ij},$$
where $m_i$, $r_i$ and $v_i$ are the mass, position, and velocity of bead $i$. The
DPD potential consists of three pairwise forces between DPD beads, *i.e.*, the conservative,
the dissipative, and the random force. The force acting on a bead is given by
$$f_{ij} = f_{ij}^c +$$
All forces vanish beyond a certain cutoff radius, $r_c$. The conservative force $F_{ij}^c$ determines the thermodynamics of the DPD system, and is defined by a purely repulsive (parabolic) soft-core potential, given by $F_{ij}^c = a_{ij}(1-r_{ij}/r_c)\tilde{r}_{ij}$, $r_{ij} < r_c$. $a_{ij}$ is the maximum repulsion between beads $i$ and $j$; $r_{ij} = r_i - r_j$, $r_{ij} = |r_{ij}|$, $v_{ij} = v_i - v_j$, and $\tilde{r}_{ij} = r_{ij}/r_{ij}$ is the unit vector directed along $j$ to $i$. The dissipative force is $F_{ij}^d = -\gamma w^D(r_{ij})(\tilde{r}_{ij} \cdot v_{ij})\tilde{r}_{ij}$ and the random force is $F_{ij}^r = \sigma_D w^R(r_{ij})\theta_{ij} \Delta t \frac{1}{2\pi} e^{-r_{ij}}$, where coefficients $\gamma$ and $\sigma_D$ characterize the strengths of the dissipative and random forces, and $\gamma = \frac{\sigma_D^2}{2k_B\tau} = 4.5$. $\theta_{ij}(t)$ is a zero-mean symmetric random variable. We use $w^D(r_{ij}) = [w^R(r_{ij})]^2 = (1 - r_{ij}/r_c)^2$ to ensure momentum conservation. Also the properties of the system are expressed using dimensionless quantities in units of the cutoff $r_c$, the energy scale $k_BT$ and the bead mass $m_0$. As a consequence, the unit of time, $\tau$, is $\tau = \sqrt{r_c^2 m_0/k_B T}$. Time evolution of the system was calculated by the Verlet algorithm with a time step $\Delta t = 0.05\tau$, where $\tau$ is the DPD unit of time. The total bead number density in the simulation system is $\rho = 3$.

One of the major advantages of DPD is the intuitiveness and ease with which simple models for various complex fluids can be constructed by modifications of the conservative interactions (repulsive parameter $a_{ij}$) between DPD beads, although there is certain limitation for DPD fluid to exhibit a rigid thermodynamic behavior of the system. However, it was demonstrated that the coarse-grained mesoscopic model can correctly
reproduce the properties and phase behavior of a system beyond certain length and time scales.\textsuperscript{21,22}

The incorporation of electrostatic interactions into DPD was originally proposed by Groot\textsuperscript{15a}, whose model is based upon the particle-particle particle-mesh (PPPM) method with a charge distribution function adapted for the use of soft potential. The second approach for the implementations of electrostatics was developed by González-Melchor and co-workers\textsuperscript{23}, using the Ewald technique\textsuperscript{24} with a charge distribution on DPD beads. DPD is extremely efficient in comparison with molecular dynamics (MD), but the implementation of long range electrostatic interactions considerably increases the computational cost and complexity of DPD. We, thus, developed a new methodology for DPD for the modeling and simulation of polyelectrolyte systems with implicit representation of ionic strength in the explicit solvent, which can efficiently capture salt-dependent conformational features of large-scale polyelectrolyte systems in aqueous solutions. Here, the ISIS DPD model is applied to study the effect of solvent ionic strength on the self-assembly of polyelectrolyte block copolymers where the repulsive parameter between polyelectrolytes in DPD is approximated with a second virial coefficient formalism in good solvent.\textsuperscript{17} The DPD conservative force produces an equation of state (EOS) that can be expressed by \(p = \rho k_B T + \alpha a \rho^2\) (\(\alpha = 0.101 \pm 0.001\)), as the virial expansion with pressure \(p\), density \(\rho\), and repulsive parameter \(\alpha = a_{ij}\), which works very well for \(\rho \geq 3\) and \(\alpha \geq 15\).\textsuperscript{19b} Thus, a good approximation for the soft potential parameter \(a_{ij}\) that holds for sufficiently high density could be \(a_{ij} \sim \nu\), where \(\nu\) is the second virial coefficient. If \(\rho = 3\), the repulsion parameter between beads of the same type is
based on the Groot and Warren\textsuperscript{10b} scheme. In the case of polyelectrolytes in a salt-dominated environment, i.e., where the bulk concentration of added salt exceeds the concentration of counterions, the short-range repulsion between charged monomers is governed by the screened Coulomb interactions (by salt) and complemented by the non-electrostatic (excluded volume) interactions. Based on the mean-field approximation, the effective second virial coefficient calculated per monomer is then equal to\textsuperscript{8,9c}:

\[ \nu = \nu_A + \alpha^2 / c_s \]  

(8.1)

where \( \nu_A \) is the bare non-electrostatic contribution to the second virial coefficient, \( \alpha \) is the degree of ionization, and \( c_s \) is the solvent ionic strength.

Analogous to the second virial coefficient, \( \nu \), in ISIS DPD model, we represent the repulsive parameters between polyelectrolyte beads as:

\[ a_{pp} = a_{ii} + a_{elec} \]  

(8.2)

\[ a_{elec} \sim c_s^{-1} \]  

(8.3)

where \( a_{ii} = 25 \) (at \( \rho = 3 \)). Only when the solvent ionic strength is extremely high, \( a_{elec} \to 0 \), the behavior of polyelectrolytes is reduced to that of neutral polymers with the same solvation properties.

The ISIS DPD model was previously\textsuperscript{17} benchmarked on the dynamics of a single long polyelectrolyte chain (300-mer) and the self-assembly of polyelectrolyte diblock copolymers (34-mer) in aqueous solutions with variable ionic strength. The main advantages of our model are its computational speed and simplicity of implementation which enables the intensive study of the self-assembly behaviors of polyelectrolyte block copolymers at high
solvent ionic strength. The main limitation of our ISIS model is that the radial gradients in the polyelectrolyte density and mobile ion distribution are neglected due to the mean-field and local electroneutrality approximations made in the derivation of this methodology. However, this effect is negligible when the polyelectrolyte chain is in salt-dominated regime, which is the case of our study in this manuscript. This model is suitable for studies of processes and properties of various systems containing strong polyelectrolytes (DNA, RNA, poly(styrenesulfonate), etc.) in salt-dominated condition.

As shown in Figure 8.1a, each polyelectrolyte diblock copolymer chain is composed of a hydrophobic block with a degree of polymerization \( N_B = 4 \) and a hydrophilic polyelectrolyte block with the variable degree of polymerization \( N_A \), ranging from 4 to 90. In our system, the amphiphilic chains are represented as a bead-spring type particle model, where adjacent beads in the chains are connected via an extra harmonic spring: \( F_{ij}^{S} = C r_{ij} \), where the spring constant \( C \) is set to 4.0 for bonds linking hydrophobic beads and 50.0 for bonds between polyelectrolyte beads to link polymer beads together in the backbone. The choice of \( C \) does not affect the qualitative behavior of the system. The repulsive parameters for the hydrophilic-hydrophobic segment interactions and water-hydrophobic interactions were set to be 90 and 100, respectively, to obtain aggregates in water. The interaction parameter between the hydrophilic polyelectrolyte segments and water molecules were set to 26.

Figure 8.1b shows the initial snapshot of the system for \( N_A = 50 \). To illustrate the self-assembly of polyelectrolyte block copolymers, 300 chains were initially distributed randomly.
in a simulation box with periodic boundaries. We consider a dilute solution of block copolymers with a polymer volume fraction ranging from 0.0729 to 0.2015. The total number of beads in the system was 139,968. The phase diagram of the aqueous solution of polyelectrolyte diblock copolymers obtained by DPD simulations are displayed in Figure 8.S1. To reveal the effect of PDC concentration on the morphology of larger aggregates, we also performed simulations at higher copolymer concentrations, i.e., with 600 chains. Figure S2 shows the equilibrium snapshots of copolymer aggregates \( N_B = N_A = 4 \).

A considerably long simulation (\( >4 \times 10^6 \) steps) was performed for each group to attain thermodynamic equilibrium. For example, the time evolution of aggregation number shows convergence by 3.0E6 steps (Figure 8.S3). After equilibrium, we have continually observed frequent exchanges of chains between micelles, and the outcome of the kinetic process guarantees true equilibrium. The trajectories were collected every 1,000 time steps. Only the last 1.0E6 time steps of the trajectories from each run were considered for the statistical analysis.

In the framework of ISIS DPD model we obtained its equilibrium characteristics as a function of the length of the hydrophilic block and ionic strength in the solution. We specifically focus here on the experimentally measurable properties of micelles: the aggregation number \( P \) (the number of chains associated into one micelle), the radius of gyration of the micelle \( R_{g,m} \) and the core \( R_{g,c} \), the thickness of the corona \( H_{corona} \), and the anisotropy \( \kappa^2 \). In some cases, multidisperse micelles were obtained in the system, especially for aspherical micelle systems. For example, the probability distribution of the
aggregation number for $N_A = 50$ is shown in Figure 8.S4. At high solvent ionic strength $a_{pp} = 30$, aggregation number displays a wider distribution for cylindrical micelles obtained. As $a_{pp}$ increased, the distribution of $P$ became narrower for spherical micelles. The size parameters for multidisperse micelles were averaged for aggregates with the aggregation number $P > 5$.

After identifying all the aggregates, our in-house script then calculated the radius of gyration of micelles and micellar cores as the root mean square average radial distance, or distance of every bead from the center of mass of the micelle, as shown by the following equation:

$$R_{g,m} = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (r_i - r_{com})^2}$$

and

$$R_{g,c} = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (r_{c,i} - r_{com})^2}.$$ 

The thickness of the micellar corona was determined using $H_{corona} = R_{g,m} - R_{g,c}$.

The script then constructed the gyration tensor for each micelle and determined the principal moments. Using the principal moments, the calculated radius of gyration could be confirmed and the relative shape anisotropy could be calculated with the following equations:

The radius of gyration was calculated as:

$$R_g^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2$$  \hspace{1cm} (8.4)

The relative shape anisotropy $\kappa^2$ was calculated as:

$$\kappa^2 = \frac{3}{2} \frac{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}{(\lambda_x^2 + \lambda_y^2 + \lambda_z^2)^2} - \frac{1}{2}$$ \hspace{1cm} (8.5)

$\kappa^2$ is bounded between 0 and 1. $\kappa^2 = 0$ only occurs if all points are spherically symmetric, and $\kappa^2 = 1$ only occurs if all points lie on a line.
The properties of DPD system are expressed using dimensionless quantities in units of the cutoff $r_c$, the energy scale $k_B T$ and the bead mass $m_0$. As a result, the unit of time, $\tau$, is $\tau = \sqrt{r_c^2 m_0 / k_B T}$. $k_B = 1.381 \times 10^{-23}$ J/K is the Boltzmann constant, and $T$ is the reference temperature in Kelvin.

In order to map the DPD model to the self-assembly system of single stranded DNA (ssDNA) amphiphiles, the above simulation parameters can be related to physical length and time scales by examining the properties of a single polyelectrolyte model in aqueous solutions.

The averaged bond length between successive beads from simulations of a single polyelectrolyte in aqueous solution is $0.2748 \pm 0.09913$ when $a_{elec} = 15$. The average segment length of polyT is approximately 6.5 Å according to the atomistic simulations of single stranded DNA [H. S. Kim et.al., In preparation], when the salt concentration is 0.5 M. This value can be used to establish a dimensional length scale in the DPD simulations and gives $r_c = 2.3$ nm. When we map one DPD polyelectrolyte bead with one monomer in polyT, the mass of one monomer is what the mass unit $m_0$ represented, which is $5.35 \times 10^{-22}$ g. As a result, $\tau = 0.26 \times 10^{-10}$ s.

8.2.2 Synthesis of Oligonucleotide Amphiphilies T50-Fluorescein:

The reaction mixtures consisted of 1 µM oligonucleotide initiator, T50, Fluorescein-dUTP (F-dUTP) monomer, (10 µM, for M:I ratios 10, respectively), and 800U of TdT in 1600 µl of terminal deoxynucleotidyl transferase buffer. Enzymatic polymerization was carried out for overnight at 37 °C after which the reaction was terminated by heating at 70 °C.
for 10 min. The reaction products were purified by centrifugal ultrafiltration (Microcon YM-10 centrifugal filter device, Millipore), followed with dialysis in Milli-Q H2O (Thermo Scientific Slide-A-Lyzer MINI Dialysis Devices, 2K MWCO) for two days, to remove, unreacted monomers, and salt compounds.

Static light scattering (SLS) measurements were performed using the ALV/CGS-3 goniometer system (ALV, Langen, Germany). Samples were prepared by filtering solution through an Anotop 10 Watman 200 nm filter into a 10 mm disposable borosilicate glass tube (Fischer). The SLS measurements were performed over a range of angles (30° to 150°) at 5° increments. Every measurement is established for 5 acquisitions with 15 seconds at each angle. A Zimm plot was created by measuring the normalized intensity of scattered light at multiple scattering angles. After linear fitting of the equation the $R_{g,m}$ was calculated from the slope.

$$\frac{\kappa c}{R} = \frac{1}{M_W} + \frac{1}{3 M_W} q^2 R_{g,m}^2$$  \hspace{1cm} (8.6)

Where $M_W$ is the weight-averaged molar mass, $R$ is the Rayleigh ratio, $q$ is the scattering wave vector, $c$ is the concentration of the sample, and $K$ is the optical constant.

**8.2.3 AFM Imaging**

Samples for AFM imaging were prepared by first placing a drop of the sample solution (a mixture of polynucleotides (~0.5 μM) in a range of NaCl concentrations with 5 mM MgCl2 added) onto freshly cleaved mica surfaces and incubating for 5 minutes. Then the sample was rinsed with Milli-Q H2O and dried in a stream of dry nitrogen. TappingMode AFM images were acquired under ambient conditions with a MultiMode AFM (Bruker),
using TappingMode silicon cantilevers \((k_F = 40 \text{ N/m}, f_{\text{res}} = 311–357 \text{ kHz}, R_{\text{tip}} < 10 \text{ nm}, \text{Bruker})\).

### 8.3 Results and discussion

We systematically analyze how the solvent ionic strength \(a_{pp}\) of the solution, and the length of the hydrophilic polyelectrolyte block \(N_A\) affect the kinetics of self-assembly and the equilibrium morphology of the aggregates (Fig. 8.1 and 8.S1). The representative aggregates of PDCs obtained by DPD simulations are displayed in Figure 8.1e and the morphological phase diagrams of final snapshots of PDCs in aqueous solutions are displayed in Figure 8.S1. Examination of the micellar shapes in Figure 1e reveals that PDCs form three morphological types, i.e., vesicle/lamellar aggregates, wormlike/cylindrical micelles, and spherical micelles. Increasing the repulsive parameter \(a_{pp}\) \(i.e.,\) decreasing solvent ionic strength \(c_s\) \) or increasing the length of the polyelectrolyte block, \(N_A\), leads to a morphological transition from vesicle/lamellar aggregates to cylindrical and spherical micelles. The changes in micellar shape are generally governed by the free energy that arises from the electrostatic or/and steric repulsion between the polyelectrolyte corona blocks and the excess free energy at the core-water interface.\(^8\, ^{9c}\) Increasing the solvent ionic strength, or decreasing the length of the polyelectrolyte blocks, lowers the repulsive interactions between the polyelectrolyte chains of the corona, which increases the size of the micellar aggregate (Figure 8.2c). Such an increase in the aggregate size implies the stretching of the core-forming chains in the radial direction\(^9c\) and leads to a conformational entropy loss in the core-
forming blocks. The morphological transformation from spherical to cylindrical micelles, and further to lamella or vesicles, leads to further relaxation of the core blocks, which is energetically favored. The influence of the core-forming blocks on the overall aggregate morphology becomes more pronounced when the size of the core exceeds that of the corona, or at high ionic strengths, as indicated at the bottom and the left side of the phase diagram. Our morphology phase diagram is consistent with the one proposed by Borisov and Zhulina’s theory for PDCs with a very large hydrophobic blocks ($N_B = 800$) and variable polyelectrolyte blocks ($N_A = 0\text{--}200$) as a function of the effective second virial coefficient (excluded-volume parameter).\textsuperscript{8,9c} Our simulations show that even with a shorter hydrophobic block ($N_B = 4$), a similar morphological phase diagram can be obtained, which indicates that the morphological diagram presented in this paper may be generally valid.

The micellar morphology can be examined in terms of the anisotropy index $\kappa^2$, where $\kappa^2$ value lies between 0 (completely spherical) and 1 (line) (see Experimental Section). Figure 8.2a shows the contour plot of the anisotropy index, $\kappa_c^2$, for the micellar hydrophobic cores. In our study we define spherical micelles as micelles with $\kappa_c^2 \leq 0.15$, whereas micelles with $\kappa_c^2 > 0.15$ are wormlike or cylindrical. Figure 8.2a shows that as $a_{pp}$ or $N_A$ increases in the wormlike/cylindrical region, $\kappa_c^2$ becomes smaller, indicating that the core approaches a more spherical shape. However, the analysis of the change in the overall micellar anisotropy, $\kappa_m^2$, in Figure 8.2b, indicates that the corona polyelectrolyte shell influences the shape of the micelles and that the $\kappa_m^2$ is quantitatively smaller than $\kappa_c^2$. In addition, an increase in $\kappa_m^2$, which occurs with long polyelectrolyte blocks at low ionic
strengths (upper right hand corner of Figure 8.2b), reflects the formation of star-like micelles, as shown in the upper-right region in the morphological phase diagram (Figure 8.1e).

The equilibrium micellar aggregation number is determined by a free energy balance between the core and corona.\textsuperscript{25} Figure 8.2c shows a contour plot of the of the averaged aggregation number, $P$, as a function of polyelectrolyte repulsive parameter, $a_{pp}$, and polyelectrolyte length, $N_A$. We observed that the average aggregation number decreases with increasing strength of the repulsive interactions, $a_{pp}$, \emph{i.e.}, decreasing solvent ionic strength. Changing the hydrophilic polyelectrolyte block length from 4 to 90, decreases the average aggregation number, which is also consistent with trends observed for neutral block copolymers.\textsuperscript{26}

The average size of a micelle is determined by an interplay between an increase in the aggregation number and polyelectrolyte chain relaxation due to electrostatic screening. When the polyelectrolyte block is short ($N_A < 30$), the micellar radius of gyration, $R_{g,m}$, decreases as $a_{pp}$ increases due to an increase in the aggregation number and the shape transitioning from cylindrical to spherical micelles (Figure 8.5). However, for a relatively long polyelectrolyte block ($N_A \geq 30$), an increase in the $R_{g,m}$ of spherical micelles is observed as $a_{pp}$ increases due to the extension of micellar corona caused by electrostatic repulsion within and between polyelectrolyte chains.

Our model allows us to derive scaling functions between different micellar parameters for spherical micelles. The contour plot of the ratio of $H_{corona}$, defined as
\[ H_{\text{corona}} = R_{g,m} - R_{g,c} \]
to the core radius \( R_{g,c} \) (Fig. 8.56) indicates that the observed spherical aggregates represent an intermediate state between star-like \( (H_{\text{corona}} \gg R_{g,c}) \) and crew-cut \( (H_{\text{corona}} \ll R_{g,c}) \) micelles. For spherical micelles, scaling functions between micellar size parameters \( (H_{\text{corona}}, R_{g,c}, \text{and } R_{g,m}) \), the repulsive parameter \( a_{pp} \), solvent ionic strength \( c_s \), (where \( a_{pp} = a_{ii} + a_{elec} \) and \( a_{elec} \sim c_s^{-1} \)), or the length of the polyelectrolyte block \( N_A \) can be used to predict the parameters of other micelles not pictured on the morphological diagram (Figure 8.3). The derived scaling exponents associated with Formula (7)-(12) are listed in Tables 8.1 and 8.2.

For all spherical micelles, we determined the scaling relations between the corona thickness, and the solvent ionic strength and the length of polyelectrolyte block (Figure 8.3a):

\[
H_{\text{corona}} \sim c_s^{-0.11} N_A^{0.75} \quad \text{(8.7)}
\]

\[
H_{\text{corona}} \sim a_{pp}^{0.2} \quad \text{(8.8)}
\]

Our results indicate that an increase in \( a_{pp} \) or \( N_A \) leads to an increase in the corona thickness \( (H_{\text{corona}}) \) despite the reduction in the aggregation number; hence, the length increase of the polyelectrolyte chains due to stretching, has a greater influence on corona thickness than the reduction of the aggregation number. Borisov and Zhulina\(^8\) obtained a scaling relationship for the thickness of corona in the crew-cut limit, \( H_{\text{corona}} \sim N_A^{0.8} v^{0.2} \), where \( v = v_A + \frac{1}{2c_s} \) is the second virial coefficient for highly charged polyelectrolyte block. In our model, the repulsive parameter between charged DPD beads, \( a_{pp} \), is the analog to \( v \) in the Borisov and Zhulina model. Thus, our scaling exponents for \( a_{pp} \) or \( N_A \) and \( H_{\text{corona}} \) are in
good agreement with the theory. Moreover, the observed scaling relation for $H_{corona} \sim c_s^{0.11}$ is in general agreement with the experimental observations. For example, Förster et al. reported that at salt concentrations above 0.05M, i.e., in the salt-dominated regime, $H_{corona} \sim c_s^{-0.13}$ for poly(ethylethylene-b-styrenesulfonic acid) (PEE–PSSH) micelles. Cristobal et al. reported a weaker dependence on ionic strength $c_s^{-0.08}$ for poly(n-butyl acrylate)-b-poly(acrylic acid) (PBA3K-b-PAA12K) micelles with a weakly charged poly(acrylic acid) block.

We observed that for spherical PDC micelles, $R_{g,c}$ generally decreased with an increasing of the repulsive parameter $a_{pp}$ or $N_A$ (Fig. 8.S7). The reduction of the core size is the result of the combined effect of the reduction in the aggregation number and the compactness of the core. We note that when $a_{pp} \geq 50$, there is a slight increase in $R_{g,c}$ when $N_A$ is increased from 70 to 90 (Fig. 8.S7b). This may be caused by a significant stretching of the longer polyelectrolyte chains which reduces the compactness of the core region. The lengths of short hydrophilic blocks, however, can have a strong influence on the core size $R_{g,c}$. Interestingly, the $R_{g,c}$ values are very close when $N_A \geq 30$ for the same solvent ionic strength. In the mean-field theory for neutral diblock copolymers, the $R_{g,c}$ is strongly dependent on the hydrophilic block length; whereas in scaling theory the $R_{g,c}$ is independent on the hydrophilic block length. Our observations indicate a crossover between the mean-field theory and the scaling behavior of the micelles which is consistent with the previous studies for diblock copolymers.
The effect of ionic strength on the radius of gyration of the micelles ($R_{g,m}$) is complex for the intermediate spherical micelles in this study (Figure 8.3b). As discussed above, decreasing ionic strength leads to a decrease in the core size and an increase in the thickness of the corona of micelles. The tradeoff between these two effects leads to the relationship between $R_{g,m}$ and solvent ionic strength. For star-like micelles, $R_{g,m}$ is expected to increase weakly with decreasing ionic strength (i.e., increasing $a_{pp}$) due to the stretching of the chains in the corona, whereas for the crew-cut micelles, a decrease in core size also leads to a decrease in $R_{g,m}$, as the ionic strength decreases. For $N_A = 30$ the system behaves more like crew-cut micelles and only a very weak increase in $R_{g,m}$ is observed as $a_{pp}$ increases (Figure 8.3b1). As the hydrophilic block length increases, the system starts to behave more like star-like micelles, so a stronger dependence of $R_{g,m}$ on $a_{pp}$ or $c_s$ was observed. Overall, the observed scaling relations between the size of spherical micelles and solvent ionic strength and the length of polyelectrolyte block are:

\begin{align}
R_{g,m} &\sim a_{pp}^{0.05}, \text{ for } N_A = 30 \\
R_{g,m} &\sim a_{pp}^{0.13}, \text{ for } N_A > 30 \\
R_{g,m} &\sim c_s^{-0.03} N_A^{0.54}, \text{ for } N_A = 30 \\
R_{g,m} &\sim c_s^{-0.06} N_A^{0.54}, \text{ for } N_A > 30
\end{align}

The theoretically predicted scaling relationship in Borisov and Zhulina’s theory for star-like micelle is $R_{g,m} \sim N_A^{0.5455} \nu^{0.091}$. While the observed $R_{g,m}$ versus length of polyelectrolyte block has an excellent agreement with the theory, the exponent for $a_{pp}$
(equivalent of $v$) in our study is slightly larger than predicted for star-like micelles (Figure 8.3b1). A weak dependence of the micelle radius ($R_{g,m}$) on solvent ionic strength is noticeable at intermediate micellar morphologies, i.e., when $H_{\text{corona}}/R_{g,c}$ ranges from 0.50 to 4.22 for spherical micelles. The weak dependence of $R_h$ on ionic strength was also observed in previous experimental studies of PIB30-b-PMAA170 micelles, i.e., $R_h \sim c_s^{-0.07}$.

Also, our $H_{\text{corona}}$ and $R_{g,m}$ scaling relation are similar to that reported by Colombani et al.\textsuperscript{30} for the self-assembly of diblock copolymers poly(n-butyl acrylate)-block-poly(acrylic acid) (PnBA–PAA): $H_{\text{corona}} \sim DP_{\text{PAA}}^{0.7}$, $R_{g,m} \sim DP_{\text{PAA}}^{0.6}$, $R_h \sim DP_{\text{PAA}}^{0.5}$, where $DP_{\text{PAA}} = N_A$ is the length of the polyelectrolyte block. Therefore, the exponent of the power law behavior of micelle radius ($R_{g,m}$) on solvent ionic strength depends on the length of the ionic block for spherical micelles at an intermediate state between star-like ($H_{\text{corona}} \gg R_{g,c}$) and crew-cut ($H_{\text{corona}} \ll R_{g,c}$) micelles.

To validate the predicted phase diagram, we compare the micellar morphology formed by highly asymmetric, 54-mer diblock polyelectrolytes using simulations and experiments. For this diblock length ($N_A = 50$ and $N_B = 4$), our simulations indicate a change in micellar morphology with increasing $a_{pp}$ (i.e., decreasing solvent ionic strength) (Figure 1e). At high ionic strength the PDCs form large micelles with aspherical cores, whereas at lower solvent ionic strength, the PDCs form spherical micelles. To verify the predicted changes we used (F-dUTP)$_4$-$b$-polyT$_{50}$, a polynucleotide block copolymer, which has a hydrophilic, 50-mer polynucleotide block (polyT$_{50}$), appended with a short, hydrophobic
oligonucleotide block (F-dUTP), which contains about 4 hydrophobic unnatural nucleotides. A combination of AFM and SLS was employed to observe the changes in micelle morphology of (F-dUTP)$_4$-b-polyT$_{50}$ as a function of monovalent salt concentration (i.e. NaCl) for a range of charge ratios, $\lambda$, defined as the ratio of charges along the polynucleotide backbones, $c_p^m$, to the monovalent salt concentration, $c_s^m$ (Figure 8.4). The charge ratio $\lambda$, representing the ionic strength, is an experimental analog of the $a_{elec}$ in the DPD simulations. In Figure 8.4a, we plot the $R_{g.m}$ obtained from SLS as a function of charge ratio $\lambda$, and $R_{g.m}$ obtained from the simulation as a function of $a_{pp}$. In addition to SLS, the assemblies were also visualized by AFM tapping mode height imaging in air at two different charge ratios $\lambda$ (Figure 4b-d).$^{11a}$ We also determined the average diameters and heights of the aggregates and micelles from AFM images (Table 8.S1).

Depending on the solvent ionic strength, our results reveal two distinct regimes for self-assembly of 54-mer PDCs. The first regime occurs at high solvent ionic strength ($\lambda$=6.1E-5 and $a_{pp} \leq 30$), where the repulsion between the polyelectrolyte chains and between adjacent micelles is reduced due to screening effects, which allows for the formation of large, micellar aggregates (Figure 8.4b). Similar micellar aggregation behavior has been observed by cryo-transmission electron microscopy (cryo-TEM) for ethylethylene-b-styrenesulfonic acid (PEE–PSSH) at a high salt concentration.$^{27}$ The second regime occurs at lower salt concentration ($\lambda \geq 3.1E-3$ and $a_{pp} > 30$), where there still is significant Coulomb repulsion between charged polyelectrolyte chains, and the micelle formation is driven by the interplay between electrostatic repulsion of charged blocks, and association of
hydrophobic blocks, which results in homogeneously distributed, individual spherical micelles (Figure 8.4c,d).

To determine the effect of ionic strength on the size of the self-assembled PDC micelles in aqueous solution we compare the salt-dependent changes in the radius of gyration ($R_{g,m}$) of micelles obtained from SLS and DPD simulations. At low ionic strengths (i.e., $\lambda > 1E^{-3}$ and $\alpha_{pp} > 30$), the $R_{g,m}$ moderately decreases with increasing salt concentration due to the increased screening of electrostatic interactions between the negatively charged polyelectrolytes. This decrease in $R_{g,m}$ reflects the gradual collapse of the micellar corona with increasing ionic strength. Upon further increase in salt concentration, the $R_{g,m}$ increases sharply, which reflects micellar aggregation and perhaps a morphological transition, induced by the decrease in electrostatic repulsion (vertical dash line in Figure 8.4a). Overall, the qualitative agreement for salt-dependent changes in the size of micelles between experiment and simulations is analogous, further validating our simulation approach.

8.4 Conclusions

Here the prediction of the salt-responsive morphologies of aggregates or micelles formed by amphiphilic diblock copolymers was achieved using ISIS DPD methodology. The copolymers in this study comprise of a short hydrophobic block and a polyelectrolyte hydrophilic block of varying lengths from 4-mer to 90-mer in aqueous solutions. We found that the morphology of the self-assembled structures undergoes transitions from spherical micelles to cylindrical micelles to lamellar aggregates with increasing solvent ionic strength.
or decreasing polyelectrolyte block length. Spherically shaped micelles were obtained over a wide area in the diagram due to the short length of the hydrophobic part. Quantitative evaluation of the micelle radius of gyration, $R_{g,m}$, and corona thickness, $H_{corona}$, and their scaling law dependence on the solvent ionic strength $c_s$ or the length of the polyelectrolyte block $N_A$ were obtained. For all spherical micelles a scaling relation $H_{corona} \sim c_s^{-0.11} N_A^{0.75}$ was obtained. The length of polyelectrolytes played a role in a scaling relationship between radius of gyration and a solvent ionic strength, where for $N_A = 30$, a scaling relation of $R_{g,m} \sim c_s^{-0.03} N_A^{0.54}$ was observed, while for $N_A > 30$, $R_{g,m} \sim c_s^{-0.06} N_A^{0.54}$.

To verify the model predictions, we studied the self-assembly behavior of polyT50-b-F-dUTP in aqueous solution as a function of solvent ionic strength. The micellar dimensions, determined from AFM and SLS measurements, showed the same conformational transitions and dependence on ionic strength as the computational prediction for a 54-mer PDC. This excellent qualitative agreement between experiment and simulations further validates our ISIS DPD simulations, and suggests that the ISIS DPD method can be used as a powerful tool to guide the rational design of solvent-responsive polyelectrolyte block copolymer nanostructures.

Acknowledgements

This work was supported by the NSF DMR-1410462, NSF's Research Triangle MRSEC (DMR-1121107) and NSF CMMI-1150682. The computer support was provided by the High Performance Computing Center at North Carolina State University. The authors thank Isaac Weitzhandler for his help with SLS measurements.
Abbreviations

PDC, Polyelectrolyte diblock Copolymer; DPD, Dissipative Particle Dynamics
8.5 References


Table 8.1 Dependence of the scaling exponents for $H_{\text{corona}}$ and $R_{g,m}$ on the repulsive parameter $a_{pp}$ and solvent ionic strength $c_s$. (represented by the electrostatic repulsive parameter $a_{\text{elec}}, c_s^{-1}a_{\text{elec}}^{-1}$).

<table>
<thead>
<tr>
<th>$N_A$</th>
<th>$H_{\text{corona}}$ and $a_{pp}$</th>
<th>$R_{g,m}$ and $a_{pp}$</th>
<th>$H_{\text{corona}}$ and $c_s$</th>
<th>$R_{g,m}$ and $c_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.196</td>
<td>0.055</td>
<td>-0.111</td>
<td>-0.031</td>
</tr>
<tr>
<td>50</td>
<td>0.217</td>
<td>0.132</td>
<td>-0.123</td>
<td>-0.074</td>
</tr>
<tr>
<td>70</td>
<td>0.248</td>
<td>0.120</td>
<td>-0.112</td>
<td>-0.053</td>
</tr>
<tr>
<td>90</td>
<td>0.235</td>
<td>0.139</td>
<td>-0.107</td>
<td>-0.063</td>
</tr>
</tbody>
</table>

Table 8.2 The dependence of the scaling exponents for $R_{g,m}$ and $H_{\text{corona}}$ on the polyelectrolyte length $N_A$.

<table>
<thead>
<tr>
<th>Repulsive parameter, $a_{pp}$</th>
<th>Scaling exponential factor between $H_{\text{corona}}$ and $N_A$</th>
<th>$R_{g,m}$ and $N_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0.761</td>
<td>0.518</td>
</tr>
<tr>
<td>50</td>
<td>0.758</td>
<td>0.540</td>
</tr>
<tr>
<td>60</td>
<td>0.746</td>
<td>0.542</td>
</tr>
<tr>
<td>90</td>
<td>0.725</td>
<td>0.545</td>
</tr>
</tbody>
</table>

Table 8.31 Micellar size from AFM images at different salt concentrations.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$6.1 \times 10^{-5}$</th>
<th>$3.1 \times 10^{-3}$</th>
<th>3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter / nm</td>
<td>$25.2 \pm 8.0$</td>
<td>$22.5 \pm 9.0$</td>
<td>$28.5 \pm 6.8$</td>
</tr>
<tr>
<td>Height / nm</td>
<td>$9.5 \pm 2.7$</td>
<td>$7.3 \pm 2.0$</td>
<td>$8.0 \pm 1.8$</td>
</tr>
</tbody>
</table>
### Table 8.S2 Experimental results.

<table>
<thead>
<tr>
<th>$\lambda$ / Charge Ratio</th>
<th>NaCl / [M]</th>
<th>Rg / nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.18904</td>
<td>5.44E-04</td>
<td>20.483±0.293</td>
</tr>
<tr>
<td>0.09443</td>
<td>0.00118</td>
<td>19.973±0.415</td>
</tr>
<tr>
<td>0.02656</td>
<td>0.00439</td>
<td>19.439±0.653</td>
</tr>
<tr>
<td>0.00366</td>
<td>0.032</td>
<td>16.726±0.515</td>
</tr>
<tr>
<td>0.0016</td>
<td>0.072</td>
<td>16.989±0.496</td>
</tr>
<tr>
<td>5.93E-04</td>
<td>0.15086</td>
<td>17.19±0.397</td>
</tr>
<tr>
<td>2.63E-04</td>
<td>0.30123</td>
<td>17.003±0.311</td>
</tr>
<tr>
<td>9.85E-05</td>
<td>0.70117</td>
<td>20.023±0.643</td>
</tr>
<tr>
<td>4.82E-05</td>
<td>1.23523</td>
<td>21.378±0.844</td>
</tr>
<tr>
<td>1.99E-05</td>
<td>2.20111</td>
<td>23.101±0.633</td>
</tr>
</tbody>
</table>
Figure 8.1 (a) Representation of a PDC chain. Simulation snapshots of initial (b) and final configurations at (c) $a_{pp} = 25$ and (d) $a_{pp} = 40$ with $N_A = 50$ and $N_B = 4$. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively. (e) Phase diagrams of typical aggregates of PDCs obtained by DPD simulations. The section views for vesicles and side views for lamellar shapes are marked by the dashed circles and ellipsoids. The phase diagrams of final snapshots of PDCs in aqueous solutions are displayed in Figure 8.S1.
Figure 8.2 Properties of micellar aggregates and morphological transitions represented as contour plots for anisotropy indices of (a) the micellar hydrophobic core, $\kappa_c^2$ and (b) the whole micelle, $\kappa_m^2$, and (c) the aggregation number, $P$. The region with aggregates formed by all chains in the system is shadowed.
Figure 8.3 The thickness of corona ($H_{\text{corona}}$) of spherical PDC micelles as a function of (a1) repulsive parameter $a_{pp}$, (a2) length of polyelectrolyte block $N_A$, and (a3) $a_{elec}$. The radius of gyration ($R_{g,m}$) of micelles as a function of $a_{pp}$ (b1), $N_A$ (b2), and $a_{elec}$ (b3). Electrostatic repulsive parameter $a_{elec} \sim c_s^{-1}$. 
Figure 8.4 (a) Radius of gyration for F-dUTP-b-polyT50 micelles determined from SLS (black squares) and from DPD simulations (blue triangles) plotted as a function of the charge ratio \( \lambda \) (bottom) and repulsive parameter \( a_{pp} \) (top). The vertical dashed line indicates the predicted change in micellar morphology. (b-d) AFM Tapping Mode height images of spherical micelles and aggregates at three different charge ratios.
Figure 8.S1a Phase diagrams of dilute aqueous solution of polyelectrolyte diblock copolymers obtained by DPD simulations. Both hydrophobic and polyelectrolyte beads are displayed in (a) and only hydrophobic beads are displayed in (b). Aggregates formed by all 300 chains, including vesicles, lamellar shape or pretzel shape are outlined with a black box, wormlike/cylindrical micelles with a red box, and spherical micelles with a blue box. The section view for vesicles and side view for lamellar shapes are displayed in dash circle.
(a)
Figure 8.S2 Typical snapshots of polyelectrolyte block copolymers in aqueous solution at $N_A=4$. Total number of chains is 300 (a1, a2, a3 and a4) and 600 (b1, b2, b3 and b4).
Figure 8.S3 Temporal evolution of averaged aggregation number $P$ for (a) $N_A=30$ and (b) $N_A=50$. 
Figure 8.S4 Probability distribution of the aggregation number when $N_A=50$. 
Figure 8.5 Radius of gyration of the micelles $R_{g,m}$ from DPD simulations as a function of repulsive parameter $a_{pp}$. 
Figure 8.6 Contour plot of ratio of the thickness of corona $H_{\text{corona}}$ to the radius of core $R_{g.c}$. The region with aggregates formed by all chains is shadowed.
Figure 8.S7 The radius of gyration of micellar core as a function of $a_{pp}$ (a) and $N_A$ (b).
Chapter 9 The Effect of Ionic Strength and Polymer Concentration on Morphology of Polyelectrolyte Triblock Copolymers

9.1 Introduction

Amphiphilic polyelectrolyte block copolymers (PBCs) exhibit a wide variety of tunable self-assembled morphologies with potential applications in both academic research and industry.\(^1\) The incorporation of polyelectrolyte block gives additional versatility to the self-assembly with regards to the responsiveness to stimuli, such as ionic strength and/or pH. Exciting nanostructures were observed through the self-assembly of PBCs in selective solvents, including disks\(^2\), hamburger\(^3\), football\(^3\), toroids\(^4\), nanotubes\(^5\) and networks\(^6\).

A remarkable number of studies have been devoted to the self-assembly properties of triblock copolymers, composed of charged middle block end-capped by similar hydrophobic short blocks, i.e. BAB triblocks.\(^7\) The combination of hydrophobic and hydrophilic/ionic blocks gives rise to the formation of multicompartment micelles and/or transient networks. The BAB triblock with a long polyelectrolyte middle block associate into flowerlike micelles which are connected to each other by bridging middle-blocks, leading to a transient physical network with visco-elastic properties.\(^8\)

It has been noted that the polyelectrolyte triblock copolymers exhibit a more complicated behavior than their neutral counterparts because the association process is controlled by the competition between the hydrophobic attraction of the insoluble blocks and
the electrostatic repulsion between the charged polyelectrolyte blocks.\textsuperscript{8} It was reported that the viscosity and elasticity of the gels can be tuned by changing polymer concentration\textsuperscript{9}, block length\textsuperscript{10}, and solvent properties, such as solvent ionic strength\textsuperscript{11}, and pH\textsuperscript{12}.

To boost the development of potential applications of polyelectrolyte gels, one has to obtain a detailed understanding of the nanostructures of the transient network. However, the fine assessment of the complicated network structure is theoretically challenging\textsuperscript{13} and also limited by the experimental difficulties in characterization. One option is to apply molecular simulations to investigate the polyelectrolyte networks. A few studies are available in which the molecular simulation techniques, such as MD, MC and DFT, are used to probe the polyelectrolyte network structure in the gelation process.\textsuperscript{14} Polyelectrolyte network simulations are generally computationally demanding because of the size of the system as well as the equilibration and relaxation times of polymer segments can be large. More importantly, the modelling of complex polyelectrolyte systems in aqueous solution in computationally expensive due to the implementation of long range electrostatic interactions, even for coarse-grained models. The latter difficulty may be overcome using the implicit solvent ionic strength (ISIS) based on DPD, wherein the explicit solvent is combined with implicit solvent ionic strength representation and permits large-scale simulation of polyelectrolytes self-assembly.\textsuperscript{15} The ISIS-DPD model has been successful in the prediction of salt-induced morphological changes of polyelectrolyte diblock copolymers and their scaling relations between micellar size parameters and solvent ionic strength or block length.\textsuperscript{16}
In this paper, we investigate how the transient network from the self-assembly of BAB polyelectrolyte triblocks responds to changes in block length, polymer concentration and solvent ionic strength by ISIS DPD models. We have chosen a polymer fraction ranging from 10% to 20%, which is above the gelation point of triblock and remarkable higher than those in earlier simulation studies. In this way, we obtained insight in the network morphology changes by calculating the aggregation number, number of hydrophobic nodes, and fraction of bridge of loop chains in the network. For the first time, the morphological phase diagrams of polyelectrolyte triblocks are built as a function of polyelectrolyte block length and solvent ionic strength.

9.2 Results and Discussions

Our simulations are based on dissipative particle dynamics models, wherein triblocks are represented by bead-spring model. Every triblock block copolymer chain contains one charged middle block with the length of $N_A=4\sim90$ which is end-capped with two hydrophobic blocks with length of $N_B=4$ (Figure 9.1a). Triblock chains of different length were initially randomly placed in a cubic simulation box with periodic boundary conditions applied in all three dimensions. Within the frame of ISIS-DPD model\textsuperscript{15}, the solvent ionic strength is represented by repulsive parameters between polyelectrolyte beads $a_{pp} = a_{il} + a_{elec}$, where $a_{elec} \sim c_s^{-1}$. In our study, $a_{pp}$ was varied between 25 and 90, which corresponds to the salt-dominated regime.\textsuperscript{15}
The final morphology of triblocks obtained at the end of each simulation is displayed in Figure 9.1c-e, with polymer volume fraction equal 10%, 15%, and 20%. As shown in Figure 9.1c, when \( N_A = N_B = 4 \), we observed raspberry shape aggregates. Triblocks with \( N_A = 10 \) form either multiple-layer hamburger (\( a_{pp} = 25 \sim 50 \)) or Swiss roll morphologies (\( a_{pp} = 70 \sim 90 \)). The aggregates with lamellar domains were also predicted through the development of a mean-field model for ABA polyelectrolyte copolymers\(^{17}\). In aqueous solutions, triblocks can form flowerlike micelles (Figure 9.1b), with the hydrophobic end blocks comprising the micelle cores and the polyelectrolyte chains adopting either a loop or bridge conformation\(^{18}\). It was predicted for neutral triblock systems that the length of middle block should be 3 times longer to form network structures.\(^{19}\) With the polymer volume fraction equal 10%, the formation of micellar network occurs when \( N_A \geq 30 \) and \( a_{pp} > 40 \).

In addition, we observed the hamburger to network crossover as solvent ionic strength decreased when \( N_A = 30 \). To the best of our knowledge, this morphological cross over has not been reported before. An explanation is that the entropy may drive the transition from hamburger to network morphologies. It was reported that the conformation of triblocks is determined by the interplay between the energy grain due to the association of the hydrophobic ends, which tends to bend the polyelectrolyte block to form loops, and the entropy loss due to the electrostatic repulsions, which tends to stretch the polyelectrolyte block. The increase in solvent ionic strength screens the electrostatic repulsion and results in the formation of larger hydrophobic domains as well as a morphology crossover from micellar network to multilayer hamburger shape.
To highlight the morphological transition from micellar network to hamburger, we characterized the aggregates by calculating the percentage of bridge chains between different hydrophobic domains and the number of hydrophobic cores. (Figure 9.2a) As the solvent ionic strength decreased ($a_{pp}$ increased), a decreasing in the fraction of bridges was observed firstly, due to the increasing electrostatic repulsion between these bridges. However, when $a_{pp} \geq 45$, a further decrease in solvent ionic strength leads to an increase in the percentage of bridge chains. This implies that further increasing loop chains will lead to the unbalance of electrostatic and hydrophobic interactions because the semi-flexible character of the central block at low solvent ionic strength condition prevents the chain from adopting a looplike conformation. Therefore, triblocks can stretch chains to balance the electrostatic repulsion and the hydrophobic attraction by inserting both of their ends into different hydrophobic cores to adopt bridge conformation. At the same time, the number of hydrophobic cores increased with morphological transition as solvent ionic strength decreased. Figure 9.2b shows the hydrophobic cores become smaller in size and approach spherical shape as solvent ionic strength decreases.

When $N_A > 30$, when the solvent ionic strength decreased ($a_{pp}$ increased), the transition from linear wormlike micelle to multicompartment micelles, e.g., segmented-worm micelles, and further to flower-like micellar network was observed. The morphologies of micelles and hydrophobic cores are displayed as an example in Figure 9.S1, for $N_A = 50$. The increase of solvent ionic strength induces the deswelling or collapse of polymer network,
which was also observed by experimental study of Polystyrene (PS) end-capped polyelectrolyte triblocks \(^{20}\) and other theoretical studies \(^{21}\).

Furthermore, we observed a similar behavior at higher polymer fraction. The morphological diagrams for systems are displayed in Figure 9.1d-e with volume fractions of 15\% and 20\% when \(N_A \geq 30\). The aggregates with helical core and connected flowerlike micelles network can be seen. Analogous morphologies for the former case have been reported for linear ABC self-assembled morphologies in middleblock-selective solvent.\(^{22}\) These results indicate that a wide range of complex structures from the self-assembly of polyelectrolyte triblocks emerge by varying the size of the blocks, solvent ionic strength and polymer volume fraction.

The aggregation number, which is defined as the number of hydrophobic tails in the micellar core, is plotted as a function of \(a_{pp}\) in Figure 9.3 and Figure 9.S2 for 10\%, 15\% and 20\% polymer volume fraction. We found that the averaged aggregation number increased gradually with increasing ionic strength when \(a_{pp} > 40\), but it increased more significantly at higher solvent ionic strength due to the morphological transition. In addition, the aggregation number decreased when the polyelectrolyte block length increased. We attribute this to the fact that the increase of the polyelectrolyte block length induces an increase of the charged corona which limits the aggregation and leads to a decrease in the aggregate size. Interestingly, at relatively low ionic strength, the aggregation number is independent on polymer volume fractions (Figure 9.S2). Figure 9.S3 shows the number of cores plotted against the repulsive parameter at varying polymer concentrations. Increasing the polymer
volume fraction or decreasing solvent ionic strength leads to the increase in the number of cores, indicating that the spatial heterogeneity diminishes.

Aforementioned, when \(N_A \geq 30\) and \(N_B = 4\), triblock formed networks of spherical micelles at low solvent ionic strength, while middle block chains may choose either a loop conformation or a bridge conformation. Bridges between the micelles in the network are considered to be the elastically active chains, the number of which is proportional to elastic modulus of polymer networks. However, it is usually difficult to determine bridge fraction directly from experimental studies; only a few experimental challenges were made to evaluate it. Our model allows us to derive scaling functions between different parameters for micellar network systems. We determined the scaling relations between the fraction of bridge chains and the solvent ionic strength at varying polymer volume fraction and polyelectrolyte block length (Figure 9.4a-c):

\[
f_{bridge} \sim c_s^\xi
\]  

(9.1)

The scaling exponent factor \(\xi\) was plotted as a function of polyelectrolyte length at different volume fractions, as shown in Figure 9.4d. Our results indicate that an increase in \(a_{pp}\) or \(N_A\) leads to an increase in the fraction of bridge chains. As we have discussed above, at low solvent ionic strength, the association process is governed by the competition between the attractive hydrophobic interactions of end blocks and the repulsive electrostatic interactions along the middle block, which leads to the formation of networks of bridged flower-like micelles. In the case of low solvent ionic strength, the looping is not favored since repulsive interactions along the chain prevent back-folding. In addition, when the
middle block is long, the back-folding of the chain is energetically unfavorable which will also results in a larger fraction of bridge chains. The fraction of bridge chains also increased as increasing polymer volume fraction due to a decrease in the inter-micellar separation. This polymer concentration effect may attribute to the excluded volume interactions\textsuperscript{25} in the polymer phase. Bridging configuration becomes more populated as the polymer volume fraction is increased, which is consistent with observations in ABC terpolymer polyelectrolyte\textsuperscript{26}.

We also observed that the scaling exponent factor generally increased with an increasing of polyelectrolyte length, however, this dependence became weaker as polymer volume fraction increased. These observations indicate that the effect of varying polyelectrolyte block length on the fraction of bridge chains becomes less significant with increasing polymer volume fraction and vice versa.

\textbf{9.3 Materials and Methods}

All simulations were performed using Dissipative Particle Dynamics (DPD) via LAMMPS\textsuperscript{27}. DPD is a coarse-grained simulation technique in which one DPD bead represents a group of atoms or a volume of fluid that is large on the atomistic scale but still macroscopically small.\textsuperscript{28} All beads move according to Newton’s equations of motion, 
\[ m_i \frac{dv_i}{dt} = \sum_{j \neq i} f_{ij}, \]
where \( m_i \), \( r_i \) and \( v_i \) are the mass, position, and velocity of bead \( i \). The DPD potential consists of three pairwise forces between DPD beads, \textit{i.e.}, the conservative, the dissipative, and the random force. The force acting on a bead is given by 
\[ f_{ij} = f_{ij}^C + \]

F_{ij}^D + F_{ij}^R. All forces vanish beyond a certain cutoff radius, r_c. The conservative force F_{ij}^C determines the thermodynamics of the DPD system, and is defined by a purely repulsive (parabolic) soft-core potential, given by F_{ij}^C = a_{ij}(1 - r_{ij}/r_c)\bar r_{ij}, \ r_{ij} < r_c. a_{ij} is the maximum repulsion between beads i and j; r_{ij} = r_i - r_j, r_{ij} = |r_{ij}|, v_{ij} = v_i - v_j, and \bar r_{ij} = r_{ij}/r_{ij} is the unit vector directed along j to i. The dissipative force is F_{ij}^D = -\gamma w^D(r_{ij})\bar r_{ij} \cdot v_{ij} and the random force is F_{ij}^R = \sigma_D w^R(r_{ij})\theta_{ij}\Delta t\bar r_{ij}, where coefficients \gamma and \sigma_D characterize the strengths of the dissipative and random forces, and \gamma = \frac{\sigma_D^2}{2 k_B \tau} = 4.5. \theta_{ij}(t) is a zero-mean symmetric random variable. We use w^D(r_{ij}) = [w^R(r_{ij})] = (1 - r_{ij}/r_c)^2 to ensure momentum conservation. Also the properties of the system are expressed using dimensionless quantities in units of the cutoff r_c, the energy scale k_B T and the bead mass m_0.

As a consequence, the unit of time, \tau, is \tau = \sqrt{r_c^2 m_0 / k_B T}. Time evolution of the system was calculated by the Verlet algorithm with a time step \Delta t = 0.05\tau, where \tau is the DPD unit of time. The total bead number density in the simulation system is \rho = 3.

In ISIS DPD model\textsuperscript{15}, we represent the repulsive parameters between polyelectrolyte beads as:

\[ a_{pp} = a_{ii} + a_{elec} \]  \ (9.2)

\[ a_{elec} = c_s^{-1} \]  \ (9.3)
where $a_{li} = 25$ (at $\rho = 3$). Only when the solvent ionic strength is extremely high, $a_{elec} \rightarrow 0$, the behavior of polyelectrolytes is reduced to that of neutral polymers with the same solvation properties.

As shown in Figure 9.1a, each polyelectrolyte diblock copolymer chain is composed of a hydrophobic block with a degree of polymerization $N_p = 4$ and a hydrophilic polyelectrolyte block with $N_A$, ranging from 4 to 90. The polymer chains were represented as a bead-spring type particle model, where adjacent beads in the chains were connected via an extra harmonic spring: $F_{ij} = C r_{ij}$, where the spring constant $C$ was set to 4.0 for bonds linking hydrophobic beads and 50.0 for bonds between polyelectrolyte beads to link polymer beads together. The choice of $C$ does not affect the qualitative behavior of the system. The repulsive parameters for the hydrophilic-hydrophobic segment interactions and water-hydrophobic interactions were set to be 90 and 100, respectively. The interaction parameter between the hydrophilic polyelectrolyte segments and water molecules were set to 26.

We consider a semi-dilute solution of triblocks with a polymer volume fraction ranging from 10% to 20%. It was reported that the triblock and the hepta-block copolymers formed free-standing gels at a concentration of 5 wt% as observed from the tube inversion test. Therefore, we assume all the simulation systems are above the gelation point. The total number of beads in the system was 139,968. A considerably long simulation ($>6 \times 10^6$ steps) was performed for each group to attain thermodynamic equilibrium. The trajectories were collected every 1,000 time steps. Only the last 1.0E6 time steps of the trajectories from each run were considered for the statistical analysis.
9.4 Conclusion

We studied the influence of solvent ionic strength and polymer concentration on the morphology of BAB polyelectrolyte triblocks, in which the A blocks carry either positive or negative charges, with the polymer volume fraction above the gelation point. Based on the simulation results, the morphological phase diagrams were built as a function of repulsive parameters between polyelectrolyte beads and polyelectrolyte block length. As the solvent ionic strength increased, the polymer network shrunk considerably leading to a morphological transition from micellar network to worm-like or hamburger-shape aggregates.

We found that an increase in ionic strength or a decrease in polyelectrolyte block length led to an increase in aggregation number. However, the aggregation number was independent of polymer volume fraction at low solvent ionic strength, even though the number of cores increased as polymer volume fraction increased. The scaling law dependence of the fraction of bridge chains on the solvent ionic strength was also obtained. The scaling exponent factor generally increased as increasing polyelectrolyte block length or polymer volume fraction. However, varying polyelectrolyte block length became less effective on the fraction of bridges with higher polymer volume fraction and vice versa.

To our knowledge, this is the first paper presenting a thorough investigation of the influence of solvent ionic strength on the self-assembly of polyelectrolyte triblocks. This study illustrates the potential of using solvent ionic strength, block length, and polymer concentration to manipulate both the morphology and the physical properties of triblock gels for many potential applications.
9.5 References


Figure 9.1 (a) Representation of a polyelectrolyte triblock chain. (b) Schematic illustration of typical loop and bridge configurations for middle block. (c-e) The phase diagrams of final snapshots of PDCs in aqueous solutions obtained by DPD simulations at the polymer volume fraction of 10%, 15% and 20%, respectively. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively.
Figure 9.2 The percentage of bridge chains (a) and the number of hydrophobic cores (b) as a function repulsive parameter $a_{pp}$ for $N_A=30$ and $N_B=4$ at polymer volume fraction of 10%.
Figure 9.3 The averaged aggregation number as a function of repulsive parameter $a_{pp}$ at polymer volume fraction of 10%.
Figure 9.4 (a-c) The percentage of bridge chains as a function of $\alpha_{elec}$ at the polymer volume fraction of 10%, 15% and 20%, respectively. (d) The exponent factor $\xi$ was plotted as a function of polyelectrolyte length.
Figure 9.S1 Different morphologies obtained with variation of the repulsive parameter $a_{pp}$ for $N_A=50$ and $N_B=4$ at polymer volume fraction of 10%. The cyan and yellow beads represent the hydrophilic and hydrophobic block, respectively.
Figure 9.S2 The averaged aggregation number as a function of repulsive parameter $a_{pp}$ for (a) $N_A=50$, (b) $N_A=70$, and (c) $N_A=90$. 

262
Figure 9.S3 The number of hydrophobic cores as a function of repulsive parameter $a_{pp}$. 
Chapter 10 Enzymatic Polymerization of High Molecular Weight DNA Amphiphiles that Self-Assemble into Star-like Micelles

Lei Tang, Vinalia Tjong, Nan Li, Yaroslava G. Yingling, Ashutosh Chilkoti, and Stefan Zauscher

Published in Advanced Materials 26 (19), 3050-3054

10.1 Introduction

Over the last few decades, the increasing insight into the structure and functionality of DNA has expanded DNA’s potential beyond that of molecular biology, and has led to the development of structural DNA nanotechnology,\(^1\) in which DNA not only encodes biological information but also serves as a polymeric, structural material that—by means of its molecular recognition properties—can be manipulated to induce self-assembly into higher order structures, such as aptamers,\(^2\) DNA origami,\(^3\) or molecular beacons.\(^4\) DNA has also been used to create hybrid amphiphiles\(^5\) that to date have been composed of hydrophilic oligonucleotides and synthetic, hydrophobic polymers.\(^1b, c, 6\) The synthesis of multifunctional polynucleotide amphiphiles with unique structural and self-assembly properties has, however, large potential for nanobiotechnology applications ranging from DNA sensing applications, nanocarriers, to drug delivery vehicles.\(^7\)
Several strategies, such as polymerase chain reaction (PCR), enzymatic ligation and rolling circle amplification, have been employed to synthesize high molecular weight (MW) DNA; however, in contrast to our approach, all these approaches require a template. Furthermore, in order to achieve amphiphilicity, conjugation with a hydrophobic, synthetic polymer has been necessary. Here, we report on a simple, biomimetic synthesis approach that harnesses an enzyme-catalyzed polymerization reaction to directly synthesize high MW, single-stranded DNA (ssDNA) amphiphiles in solution and with low polydispersity. Specifically, we use the template-independent polymerase terminal deoxynucleotidyl transferase (TdT), which sequentially adds 3’-deoxyribonucleoside 5’-triphosphates (dNTPs) to an oligonucleotide primer. This enzyme-catalyzed polymerization of DNA is conceptually similar to controlled or “living” polymerization of synthetic polymers, in which monomers are added to a growing chain one at a time. To the best of our knowledge, this template-independent enzymatic polymerization approach has not yet been used for the synthesis of high MW ssDNA amphiphiles and copolymers.

10.2 Experimental Section

Synthesis of homopolynucleotides (poly(dTTP)) with three different MWs (designated as pT1, pT2 and pT3) were accomplished by enzymatic polymerization using three different M:I ratios (200, 500, and 1000). The reaction mixtures consisted of 500 μM dTTP monomer, Cy5 (Cyanine 5 with Absorbance Max at 648 nm and Emission Max at 668nm) labelled oligonucleotide initiator, i.e., 5’-Cy5-dT10, (2.5 μM, 1 μM, and 0.5 μM, for M:I ratios 200,
500, and 1000, respectively), and 200U of TdT in 200 μl of TdT buffer. Enzymatic polymerization was carried out for 2 h at 37°C after which the reaction was terminated by heating at 70°C for 10 min. The reaction products were purified by centrifugal filtration (Microcon YM-30 centrifugal filter device, Millipore), to remove non-extended initiators, unreacted monomers, and denatured TdT subunits. The concentration of purified homopolynucleotides was determined from the Cy5 fluorescence of the initiator using a Nanodrop fluorimeter (Thermo Scientific). The percentages of the extended initiators (i.e. home-polynucleotides) to initiators are 78 ± 0.5 %, 78 ± 0.3 %, and 77 ± 0.9 %, for pT1, pT2 and pT3, respectively. The hydrophobic BODIPY-dUTP tail was appended to the hydrophilic poly(dTTP) block again by enzymatic polymerization using TdT, with a 10:1 excess of BODIPY-dUTP monomer. Here, the reaction mixtures consisted of 0.5 μM poly(dTTP), serving as the Cy5 labelled macroinitiator, 5 μM BODIPY-dUTP monomers and 100U of TdT in 100 μl of TdT buffer. Enzymatic polymerization was carried out overnight at 37°C after which the reaction was terminated by heating at 70°C for 10 min. The reaction products were again purified by centrifugal filtration (Microcon YM-30 centrifugal filter device, Millipore).

The number of BODIPY moieties per polynucleotide chain was determined by BODIPY fluorescence measurements after enzymatic treatment with exonuclease I (BioLabs). Exonuclease I catalyzes the removal of BODIPY-dUTP mononucleotides from ssDNA in the 3’ to 5’ direction. This fragmentation into individual BODIPY-dUTP
nucleotides prevents BODIPY self-quenching, and allows for quantitative fluorescence measurements (see SI for more details).

*Samples for AFM imaging* were prepared by placing a drop of the sample solution (~0.5 μM polynucleotide amphiphiles in a mixture of 300 mM NaCl and 5 mM MgCl₂) onto freshly cleaved mica surfaces. After 5 min incubation the samples were rinsed with Milli-Q H₂O and then dried with N₂. TappingMode AFM images were acquired under ambient conditions with a MultiMode AFM (Bruker), using TappingMode silicon cantilevers (kᵥ = 40 N/m, fᵥres = 311-357 kHz, RTip < 10 nm, Bruker).

*DLS (DynaPro, Wyatt Technology) experiments* were carried out at 25°C using a single detector at 90°. All samples were filtered through 0.2 μm syringe filters (Anotop 10, Whatman) before measurements. The data was analyzed with a regularization fit (multimodal) provided by Wyatt Technology.

*DPD simulations* were carried out with the LAMMPS program²⁰ for a random dispersion of the long amphiphile chains in solution in a cubic cell of size 40×40×40 rᵥc³, where rᵥc is the DPD unit of length. To investigate the effect of polynucleotide block length on assembly behaviour, we varied the hydrophilic block length, NA = 300, 500, 600 and 900, while keeping the hydrophobic block length the same, NB = 4. The volume fraction of amphiphiles was 0.0158, 0.0262, 0.0315, and 0.047 for polynucleotide block lengths of 300, 500, 600 and 900, respectively. The repulsive force parameters between the same bead type were equal and set to aᵢᵢ = 25 kᵥB T and the bead density of the system corresponded to a liquid at room temperature. The negative charges of the polynucleotide chain segments were
considered to be screened, reflecting the high ionic strength used in the experiments. The repulsive parameters between the hydrophilic-hydrophobic segments and between the hydrophobic segments and water (solvent) were set to $200k_BT$ to obtain aggregates. The parameter between the hydrophilic polynucleotide segments and water were set to $26k_BT$. To reach dilute solution conditions ten amphiphile chains were present in each simulation system. Time evolution of the system was calculated by the Verlet algorithm, with a time step of $0.05\tau$, where $\tau$ is the DPD unit of time. A considerably long simulation (> $6\times10^6$ steps) was performed for each group to attain thermodynamic equilibrium.

10.3 Results

The two-step enzymatic polymerization reaction is shown in Figure 10.1a). In the first step (1) a homo-polynucleotide (Cy5-dT$_{10}$)poly(dTTP) or poly(dTTP) in short, is enzymatically synthesized by TdT, which sequentially adds dTTP to the terminal 3’-OH group of a fluorescently-labelled primer, 5’-Cy5-dT$_{10}$, which we term the “initiator.” In the second step (2) a short sequence of an unnatural nucleotide—BODIPY-dUTP (abbreviated as B-dUTP)— is appended to the terminal 3’-OH group of the poly(dTTP), again by TdT catalysis. Although TdT can catalyze the addition of a structurally diverse set of mononucleotides from the 3’-OH group of ssDNA, we chose B-dUTP for several reasons. First, B-dUTP is a hydrophobic molecule and we aimed to incorporate multiple copies of a hydrophobic nucleotide at the terminus, so as to drive the self-assembly of the polynucleotide. We note, however, that TdT has been used previously to incorporate a single, unnatural
nucleotide at the end of a DNA for labeling DNA fragmentation,\textsuperscript{12} and alternatively has been used to incorporate multiple copies of unnatural nucleotides that are “stochastically doped” along a ssDNA chain that is largely composed of natural deoxynucleotides.\textsuperscript{13} To date, however, TdT has not been used to synthesize polynucleotide copolymer amphiphiles by appending multiple copies of a hydrophobic nucleotide to the 3′ terminus of a hydrophilic polynucleotide block. Second, BODIPY is a fluorophore, which allows us to quantify its incorporation by fluorescence spectroscopy. Third, although BODIPY is fairly bulky, we speculated that its incorporation should be feasible by TdT-catalyzed polymerization, because dUTP is modified by BODIPY at the C-5 position of the uridine base by an alkynyl-amino linker, and the catalytic activity of TdT primarily depends on the identity of the triphosphate residues of the dNTP, whereas modifications to the base and sugar moieties are less important.\textsuperscript{10b, 14}

The MW of enzymatically polymerized ssDNA is controlled by the molar ratio of monomer (M) to initiator (I), the type of the nucleotide used as the monomer, the enzyme concentration, and the reaction time.\textsuperscript{10b, 13-14} We chose dTTP as the monomer, 5′-Cy5-dT\textsubscript{10} as the initiator, and fixed the enzyme concentration (1U/μl) and reaction time (2 h). We used three M:I ratios (200:1, 500:1, and 1000:1) to synthesize homopolynucleotides (poly(dTTP)) with three different MWs, labelled pT1, pT2, and pT3 in Figure 10.2a. We determined the molecular weight of these polynucleotides by agarose gel electrophoresis (Figure 10.2). Specifically, we used the 5′-terminal Cyanine 5 (Cy5) fluorophore in the initiators to visualize the poly(dTTP) bands in the gel. The MWs of pT1, pT2 and pT3 were determined
from the peak intensity of the bands in the gel, and are 0.7, 1.6, and 2.7 kb, respectively. About 50% of the poly(dTTP)s were subsequently functionalized by appending several, unnatural, hydrophobic nucleotides (B-dUTP) to the 3'-OH terminal group, again by TdT catalyzed chain extension. These products are labelled pT1B, pT2B, and pT3B in Figure 10.2b.

The narrow bands shown in Figure 1 suggest a relatively narrow MW distribution of the poly(dTTP). To quantify the MW distributions of our polynucleotides, we determined the number average molecular weight ($M_n$), the weight average molecular weight ($M_w$), and the polydispersity index ($PDI = M_w/M_n$) from the Cy5 fluorescence distribution in a band (see SI for details). This approach yielded $PDI$ values for the poly(dTTP) homopolymers and the poly(dTTP) amphiphiles ranging from 1.03 to 1.09; i.e., $PDI$ values that approach those of synthetic polymers synthesized by controlled or “living” polymerization reactions. The analogy to living polymerization reactions is further apparent by the observation that the $PDI$ decreases with increasing molecular weight of the polynucleotide chains. However, the measured MWs are higher than those expected from the M:I ratios, which suggests an initiation efficiency of less than 100%. This effectively increases the M:I ratios, making more monomers available for the polymerization reaction.

It is not possible to resolve the length of the B-dUTP containing block by gel electrophoresis, because the MW difference is less than 1% in all cases. This means that the second, B-dUTP containing block is less than 0.1 kb in all three cases, which suggests that our diblock copolymers are highly asymmetric, as expected. The incorporation of the B-
dUTP in the second block was verified by scanning the agarose gel for the fluorescence emission of the BODIPY moieties (emission at 515 nm) (Figure 1b). While this measurement only confirms the presence of B-dUTP, it does not allow determination of the average number of appended B-dUTP, because the close proximity of BODIPY moieties to each other quenches their fluorescence intensity. To determine the average number of B-dUTP per chain by fluorescence measurements, we resolved the B-dUTP self-quenching issue by treating the amphiphiles with exonuclease I. This enzyme catalyzes the sequential removal of deoxynucleotides in the 3’ to 5’ direction and releases brightly fluorescent BODIPY-modified nucleotides into solution. We then determined their concentration by fluorescence measurements, and calculated an average between 3 to 4 BODIPY moieties per chain (3.78 ± 0.04, 3.56 ± 0.02 and 3.13 ± 0.05 for pT1B, pT2B, and pT3B, respectively). These results show that although TdT is able to incorporate a wide range of unnatural dNTPs, the incorporation efficiency for the bulky B-dUTP is low compared to that of the natural dNTPs. This likely results from steric constraints imposed by B-dUTP on TdT’s catalytic activity.

In these DNA amphiphiles, the poly(dTTP) block acts as the hydrophilic polyelectrolyte block and the much shorter BODIPY-containing block appended at the 3’ terminus of the poly(dTTP), provides the hydrophobic segment. In aqueous solution, the hydrophilic poly(dTTP) blocks are hydrated and adopt an extended conformation. Above a critical micelle concentration (CMC) of the amphiphiles, the hydrophobic BODIPY tails aggregate and form the micellar core. Because the hydrophobic block is far smaller than the hydrophilic block, the resulting self-assembled structures are “star-like” micelles with a small
hydrophobic B-dUTP core and a hydrophilic poly(dTTP) corona. This is shown schematically in Scheme 1b and was experimentally verified by atomic force microscopy (AFM) (Figure 10.3). The self-assembly of the polynucleotide amphiphiles occurred in aqueous solution at moderately high salt concentration (300 mM NaCl). At this ionic strength, electrostatic interactions are screened \((\kappa^{-1} \sim 0.5 \text{ nm})\), and the chain conformation of the hydrophilic polynucleotide block is indistinguishable from that of an equivalent neutral polymer.\(^{19}\)

To further explore the self-assembly of polynucleotide amphiphiles, we used Dissipative Particle Dynamics (DPD), implemented using the Lammps software package; details are provided in Supporting Information.\(^{20}\) DPD is a meso-scale simulation technique that is used to study and predict the phase behavior and properties of copolymer aggregates as well as complex self-assembled structures such as multicomponent micelles and polymersomes.\(^{21}\) All DPD simulations started from a random dispersion of long amphiphile chains representing ssDNA in coarse-grained water. To investigate the effect of DNA block length on self-assembly behaviour, we varied the hydrophilic block length, \(N_A = 300, 500, 600\) and 900, while fixing the hydrophobic block length at \(N_B = 4\).

Figures 10.2g) and 2h) show snapshots of the equilibrium morphology of ssDNA micelles, taken at the end of the simulation trajectory, which are consistent with the formation of star-like micelles observed by AFM. Furthermore, DPD simulations show that with increasing hydrophilic block lengths from 300 to 900 nucleotides, the aggregation number initially decreases from 9 to 5 and then stabilizes at 5 for block lengths exceeding
600 nucleotides (Table 10.S3). These computationally determined aggregation numbers agree well with those determined from AFM images for $N_A > 600$ (Table 10.1).

The AFM images (Figure 10.3) of the self-assembled micellar nanostructures show a clear distinction between a condensed core and a hairy corona. However, the diameters of the condensed cores are much larger than their heights. This arises in part from spreading of the micelles on the mica substrate during drying and likely also reflects the tip broadening effect inherent to AFM. To better approximate the micellar core size, we determined the average heights (4.0 ± 1.3 nm, 4.7 ± 1.3 nm and 6.3 ± 1.9 nm, for pT1B, pT2B and pT3B, respectively) of the condensed cores from height distributions (Figures 2b, d, f) obtained from the AFM images. We observed an approximately linear height increase of the condensed cores with increasing MW of the poly(dTTP) chains in the corona. This likely arises from the wrapping and aggregation of the corona arms around the hydrophobic micellar core, as also shown in the snapshots from the DPD simulations (Figures 10.2g, 10.2h). From the AFM images, we also estimated the aggregation number ($m_{AFM}$), i.e., the number of amphiphiles forming a micelle, by counting the number of arms emanating from the cores of the micelles. The data in Table 10.1 show that the aggregation number is essentially constant, ranging between 4-5, regardless of the length of the hydrophilic poly(dTTP) block. This observation is supported by the predictions of the DPD simulations, and is consistent with the theory for star-like micelles, developed by Halperin.22

We next determined the size of unmodified homopolynucleotides pT1, pT2, and pT3, and the size of the micelles, formed by the amphiphiles pT1B, pT2B, and pT3B, by dynamic
light scattering (DLS) in aqueous solution (300 mM NaCl). The mean hydrodynamic radii \( (R_h) \) (Table 10.1) were obtained from multimodal fits of the autocorrelation data with a regularization model. The \( R_h \) of the unmodified homopolynucleotides increased from about 9 nm to 21 nm, corresponding to their MW increase from 0.7 to 2.7 kb (Table 10.1, and Figure 10.S4). The observed \( R_h \) distributions are uniform and narrow, consistent with the narrow molecular weight distributions obtained by gel electrophoresis (Figure 10.S1). The \( R_h \) distributions of the micelles are shown in Figure 10.S4. The mean \( R_h \) of the major peaks increases from about 33 nm to 57 nm and reflects the increase in corona thickness with increasing molecular weight of the homopolynucleotide chains in the micelles. The measured \( R_h \) values are, however, smaller than the overall size of the star-like micelles, due to the low segment density in the corona as illustrated in the AFM images shown in Figure 10.3. As expected, the \( R_h \) of the micelles increases with decreasing ionic strength (Table 10.S2), consistent with an increase in electrostatic repulsion between the corona chains.

For these highly asymmetric amphiphiles, where the size of the hydrophilic block \((N_A)\) is much larger than that of the hydrophobic block \((N_B)\), and where charges are effectively screened, scaling law theory predicts that the curvature of the core becomes apparent, and that the micellar structure is best described as a “spherical quasineutral brush”—regime I(s) in Shusharina et al.\(^{23}\) In that regime, the aggregation number \((m)\) scales with \( N_B^{4/5} \), the corona thickness \((R_A)\) scales with \( N_A^{3/5}/N_B^{4/25} \), and the core radius \((R_B)\) scales with \( N_B^{3/5} \).\(^{22}\) Furthermore, our DPD simulations predict that the experimentally observed micelles should obey scaling law theory since the lengths of the hydrophilic blocks used in
the experiments all exceed 600 nucleotides. To test this scaling, we plotted the micelle radius of gyration ($R_g$) at fixed aggregation number ($m = 5$) from DPD simulations, and the $R_h$ for micelles from DLS, against the hydrophilic block length ($N_A$) at constant hydrophobic block length ($N_B = 4$). Figure 10.4 shows that the $R_g$ scaling from DPD data closely follows the theoretical prediction while the $R_h$ scaling from DLS falls slightly below the theoretical prediction of the power of $3/5$. These results suggest that both experimental and simulation results are consistent with theoretical scaling law predictions.

We determined the CMC of our polynucleotide amphiphiles from the maximum change in the concentration-dependence of the BODIPY fluorescence intensity measured in 300 mM NaCl solution by fluorescence spectroscopy. We used the Cy5 fluorescence, which does not quench with increasing concentration, as an indicator for concentration. The relative fluorescence intensities of BODIPY are plotted as a function of the concentration of amphiphilic polynucleotides (pT1B, pT2B, and pT3B) in Figure 10.S5. The CMC for pT1B, pT2B, and pT3B are 0.50, 0.53, and 0.67 μM, respectively. The increase in the hydrophilic polynucleotide chain length compared to the constant length of the hydrophobic tail, can explain the slight increase in the CMC values with increasing poly(dTTP) length.

In summary, our results are notable for several reasons. First, the synthesis of ssDNA in the kilobase range has been challenging to date. Solid phase methods which are often used can add only up to ~150 nucleotides in length with reasonable yield. TdT-catalyzed DNA polymerization, however, enables the facile synthesis of high MW ssDNA with low polydispersity. Although we can control the nucleotide sequence only to a limited extent,
TdT does not require a template to synthesize high MW block-co-polynucleotides. Second, our results demonstrate, for the first time, the synthesis of highly asymmetric, amphiphilic DNA diblock-copolymers that can self-assemble into hairy, star-like micelles. Third, the results from our experiments and DPD simulations are selfconsistent and agree with theoretical predictions. This suggests that DPD simulations could be used to predict the micellization behavior of a potentially wider range of polynucleotide copolymers, and thus guide the synthesis of useful polynucleotide architectures. Fourth, this enzymatic polymerization of natural and unnatural nucleotides into functional macromolecules with self-assembly behavior is exciting and useful for the synthesis of micellar structures that could find therapeutical applications in biomedicine. We are currently synthesizing star-like micelles that could function for drug delivery, in which the building blocks are also the drug (i.e., unnatural nucleotides, such as 5-Fluorouracil are an integral part of the polynucleotide strands), and that use targeting aptamers as initiator. Ours is the first example of high MW polynucleotide amphiphiles that consist solely of nucleotides and that display self-assembly in solution, which opens up new directions of research and potentially new applications for this new class of polymers.

Acknowledgements

SZ and AC acknowledge support through NSF CBET-1033621. YY acknowledges support through NSF CMMI-1150682. SZ, AC, and YY acknowledge support through NSF MRSEC (DMR-1121107).
Supporting Information Available: This material is available free of charge via the Internet at http://onlinelibrary.wiley.com/doi/10.1002/adma.201306049/full.
10.4 References


Table 10.1 Average hydrodynamic radii \( (R_h) \) and micellar core heights \( (H_{core}) \) obtained from DLS measurements and AFM images, respectively.

<table>
<thead>
<tr>
<th>Polynucleotides</th>
<th>Micelles</th>
<th>Polynucleotides</th>
<th>Micelles</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_h ) / nm</td>
<td>Intensity</td>
<td>( R_h ) / nm</td>
<td>Intensity</td>
</tr>
<tr>
<td>T1</td>
<td>8.8±1.0</td>
<td>93.1±5.5</td>
<td>T1B</td>
</tr>
<tr>
<td>T2</td>
<td>18.1±0.2</td>
<td>99.9±0.2</td>
<td>T2B</td>
</tr>
<tr>
<td>T3</td>
<td>21.3±0.3</td>
<td>99.7±0.4</td>
<td>T3B</td>
</tr>
</tbody>
</table>

\(^{a)}\) \( H_{core} \) is the average micellar core height determined from AFM images;

\(^{b)}\) \( m_{AFM} \) is the estimated aggregation number from AFM images.
Figure 10.1 Schematic, showing a) the enzyme-catalyzed synthesis of diblock polynucleotide amphiphiles, and b) their self-assembly into star like micelles.
Figure 10.2 Agarose gel electrophoresis results showing a) enzymatically synthesized homopolynucleotides (poly(dTTP)) with three different MWs, using M:I of 200:1, 500:1, and 1000:1 (lanes pT1-pT3), and b) the same three poly(dTTP) enzymatically end-functionalized with B-dUTP (lanes pT1B-pT3B), at a molar ratio of B-dUTP to poly(dTTP) of 10:1. The presence of B-dUTP was verified by BODIPY fluorescence.
Figure 10.3 AFM TappingMode height images of star like micelles (pT1B, pT2B and pT3B), and the corresponding histograms of the measured micellar core heights. Inset: 3-D images of typical micelles (scale bar: 50 nm XY, 1.5 nm Z) g-h) Snapshots of equilibrium ssDNA micelles from the DPD simulations trajectory, where yellow beads represent the hydrophobes and blue beads represent hydrophilic polynucleotides.
Figure 10.4 Micelle $R_g$ at fixed aggregation number ($m = 5$) from DPD simulations, and the $R_h$ for micelles from DLS, plotted against hydrophilic block length ($N_A$). The slope of the lines through the data points reflects the power in the scaling law relationships.
Supporting information for

Molecular description of the LCST behavior of an elastin-like polypeptide

Nan K. Li†, Felipe García Quiroz‡, Carol K. Hall§, Ashutosh Chilkoti‡, and Yaroslava G. Yingling*†

†Department of Materials Science and Engineering and §Department of Chemical and Biomolecular Engineering, North Carolina State University, 911 Partners Way, Raleigh, North Carolina 27695, United States

‡Department of Biomedical Engineering, Duke University, P.O. Box 90281, Durham, North Carolina 27708, United States

288
Figure 3.1. Representative structures of (VPGVG)$_{18}$ from cluster analysis of MD trajectories at (a-c) 290 K and (d-f) 350 K. The value of the non-bonded energy of the peptide is reported in Table 3.S1.
Figure 3.S2. Structure of (VPGVG)$_{18}$ at the end of a 70 ns simulation at different temperatures. The backbone is represented as secondary structure ribbons and all heavy atoms are depicted as sticks. Side chain is colored by amino acid (Gly in purple, Pro in yellow and Val in blue) and turn motif is colored in green, β-sheet is colored as yellow and coil motif is silver.
Figure 3.S3. Temporal evolution of radius of gyration ($R_g$) of $(VPGVG)_{18}$ at (a) different temperatures and (b) at 300K extended to 500 ns.
Figure 3.S4. Time autocorrelation function of peptide backbone Cα- Cα bond vectors: (a) unit vector along backbone Cα- Cα bond; (b) unit vector out of plane.

We can estimate the relaxation time of peptide backbone vectors from the time autocorrelation profiles which is less than 30ns for each group.
Figure 3.S5. Radial distribution function of oxygen atoms in water molecules around atoms in backbone (a) and side chain (b). Radial distribution function of oxygen atoms in water molecules around polypeptide atoms as a function of temperature at different time intervals. The positions of peaks are approximately the same for all temperatures.
Figure 3.S6. Probability distribution of the size $N_{max}^w$ of the largest water network in pure water system with
Figure 3.S7. Secondary structure formation as a function of temperature for turn and β-strand structures at different time intervals.
Figure 3.S8. Average frequency of occurrence of each residues in (VPGVG)$_{18}$ in each type of secondary structural motifs using DSSP program$^2$ implement by Amber Tool at 290 K (a) and 350 K (b).
Figure 3.S9. Representative structures for G(VPGVG)V segment at 290 K (a) and 350 K (b) built with the dihedral angles obtained from the highest intensity points in Ramachandran plots (Figure 3.4)
Figure 3.S10. (a) Radius of gyration $R_g$ of $(VPGVG)_{18}$ chains at 290 K and 350 K from simulations of two peptides. (b) Temporal evolution of average structural propensities over all residues for β-strand structure and turn structure.
Supporting information for

LCST Behavior is Manifested in a Single Molecule: Elastin-like polypeptide (VPGVG)$_n$

*Binwu Zhao$^†§$, Nan K. Li$^‡§$, Yaroslava G. Yingling$^‡$, and Carol K. Hall$^{*†}$

†Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27606, United States

‡Department of Material Science and Engineering, North Carolina State University, Raleigh, North Carolina 27606, United States

299
Figure 4.S1. Ramachandran plot showing peak values for ELP (VPGVG)$_n$ of 3 different lengths, L=10, 18, 30 at temperature (a) $T<T_t$ and (b) $T>T_t$. Each region has been circled with colors corresponding to residues on different positions. Ramachandran plots of L=10 and 30 have the same intensity change as (VPGVG)$_{18}$ in previous report by Li (39).
Figure. 4.S2. Probability distribution of the size $N_{\text{max}}$ of the largest water network around ELP (VPGVG) of three lengths $L=(a)$ 10, (b) 18 and (c) 30. In each figure, three representative temperatures are included, a low temperature, an intermediate temperature and a high temperature.
Figure 4.S3. Time autocorrelation function of peptide backbone Cα-Cα bond unit vectors out of plane. Relaxation time of peptide backbone vectors can be estimated from time autocorrelation profile for (A) (GVG)(VPGVG)3, (B) (VPGVG)10, (C) (VPGVG)18 and (D) (VPGVG)30 when it first reaches zero. The relaxation time is less than 35 ns for ELP (VPGVG)n of all lengths.
A $(VPGVG)_3$: 290 K, 300 K, 305 K, 310 K, 320 K.
(VPGVG)$_3$: 450 K, 470 K, 490 K.
(VPGVG)$_{10}$: 300 K, 305 K, 310 K, 320 K, 325 K.
\( (VPGVG)_{18}: \) 290 K, 300 K, 305 K, 310 K, 320 K.
(VPGVG)$_{18}$: 325 K, 330 K, 335 K, 340 K, 350 K.
D

(VPGVG)$_{30}$: 285 K, 300 K, 305 K, 310 K, 320 K.
(VPGVG)$_{30}$: 325 K, 335 K, 350 K.
Figure. 4. S4. Radial distribution function of oxygen atoms in water molecules around atoms in peptide of different lengths: (a) (GVG)(VPGVG)$_3$ (b) (VPGVG)$_{10}$ (c) (VPGVG)$_{18}$ (d) (VPGVG)$_{30}$. The position of the first minimum that each curve reaches are approximately the same for all lengths around 2.45 Å, water molecules within this distance to peptide backbone atoms are in the first hydration layer, and are used for hydration analysis.
Figure 4.S5. Secondary structure formation as a function of temperature for turn and β-strand structures at different time intervals for (A) (VPGVG)_{10} (B) (VPGVG)_{18} and (C) (VPGVG)_{30}. 
A (VPGVG)$_{10}$
B (VPGVG)$_{18}$:
$C(VPGVG)_{30}$
Figure 4.S6. Number of water molecules in the hydration shell of (VPGVG)$_{30}$ versus temperature.
Supporting information for
Enzymatic Polymerization of High Molecular
Weight DNA Amphiphiles That Self-Assemble
into Star-Like Micelles

Lei Tang, a Vinalia Tjong, b Nan Li, c Yaroslava G. Yingling, c,* Ashutosh Chilkoti, b,* and
Stefan Zauscher a,*

L. Tang, Prof. S. Zauscher
Dept. of Mechanical Engineering and Materials Science, Duke University, 144 Hudson Hall
Box 90300, Durham, North Carolina 27708, USA
E-mail: zauscher@duke.edu
Dr. V. Tjong, Prof. A. Chilkoti
Dept. of Biomedical Engineering, PO Box 90281, Duke University, Durham, North Carolina
27708, USA
E-mail: chilkoti@duke.edu
N. Li, Prof. Y. G. Yingling
Dept. of Materials Science and Engineering, North Carolina State University, Raleigh, North
Carolina 27695, USA
E-mail: yara_yingling@ncsu.edu
Materials and Methods

Materials

The oligonucleotide initiators (5'-Cy5-dT10) for this study were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). TdT enzyme, TdT buffer, and natural dNTP monomers (dTTP) were purchased from Promega (Madison, WI). ChromaTide® BODIPY® FL-14-dUTP was purchased from Life Technology (Carlsbad, CA). Top Vision agarose was purchased from Thermo Scientific (Waltham, MA). Exonuclease I (E. coli) was purchased from New England BioLabs (Ipswich, MA).

Gel Electrophoresis and Polydispersity Analysis

Gel electrophoresis was conducted on a horizontal electrophoresis system (C.B.S Scientific Company, Inc), by loading a 10 μl sample (~0.5 μM) into 2% SYBR Green II stained agarose gel, and then applying 120 V for 45 min. The gels were imaged with a Typhoon 9410 scanner (GE Healthcare Life Science, Piscataway, NJ) at 633 nm (Cy5), and 488 nm (BODIPY and SYBR Green II) laser excitation. The MWs of the samples were estimated using a 0.1-2 kb or a 0.5-10 kb RNA ladder (Invitrogen, CA).

The Cy5 scans shown in Figure 10.1 were used to determine the polydispersity of our polynucleotides, by analyzing the fluorescence intensity of each band in the gel with ImageJ (NIH). For this, a rectangular section (region of interest, ROI) was defined, stretching from the sample loading well to the edge of the gel, centered in a lane and covering 75% of the total lane width, and the gray-value of each lane was then plotted against the migration distance (Scheme S1). To establish a standard curve, the relationship between the electrophoretic mobility and the molecular weight standard from the RNA ladder was determined (Scheme S1). Using this standard curve we were then able to convert the relative mobilities of the DNA samples to the corresponding MW (in kb), as shown in Figure 10.S1, where the relative fluorescence intensity (y-axis) extracted from the ROI is plotted against MW (x-axis). The weight average (\(M_w\)) and number average (\(M_n\)) molecular weights were calculated by analysis of Figure 10.S1, using Equation 1 and 2, and the polydispersity index (PDI) was defined as \(M_w/M_n\). A similar method has been used to determine the molecular weight distribution of hyaluronan samples.[1] To determine the average molecular weights, we considered that macromolecules exist in discrete fractions \(i\) containing \(N_i\) molecules of molar mass \(M_i\).
\( \overline{M_n} = \sum x_i M_i = \sum N_i M_i / \sum N_i \) \hfill (1)

\( \overline{M_w} = \sum w_i M_i = \sum N_i M_i^2 / \sum N_i M_i \) \hfill (2)

Here \( x_i \) and \( w_i \) are the mole fraction and weight fraction of molecules of molar mass \( M_i \), respectively.

The average molecular weights and the PDI values for homopolynucleotides (pT1, pT2, and pT3) and the BODIPY modified poly(dTTP) (pT1B, pT2B, and pT3B) are listed in the Table S1.

**Scheme S1.** Relationship between the molecular weight of the RNA standard ladder and the electrophoretic mobility.

**Figure 10.S1.** Relative fluorescence intensity distribution for polynucleotide samples plotted against molecular weight.
Table 10.51. Average molecular weight and PDI values for poly(dTTP) and BODIPY modified poly(dTTP).

<table>
<thead>
<tr>
<th></th>
<th>pT1</th>
<th>pT2</th>
<th>pT3</th>
<th>pT1B</th>
<th>pT2B</th>
<th>pT3B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{M}_n$ (kb)</td>
<td>0.62</td>
<td>1.62</td>
<td>2.91</td>
<td>0.69</td>
<td>1.70</td>
<td>3.10</td>
</tr>
<tr>
<td>$\bar{M}_w$ (kb)</td>
<td>0.68</td>
<td>1.72</td>
<td>3.09</td>
<td>0.76</td>
<td>1.78</td>
<td>3.20</td>
</tr>
<tr>
<td>PDI</td>
<td>1.10</td>
<td>1.06</td>
<td>1.06</td>
<td>1.10</td>
<td>1.05</td>
<td>1.03</td>
</tr>
</tbody>
</table>

The enzyme-catalyzed polymerization is in many ways similar to controlled or “living” polymerizations of synthetic polymers, in which monomers are added to a growing chain one at a time, and is apparent by the observation that the PDI decreases with increasing molecular weight of the polynucleotide chains.

Determination of the Average Number of BODIPY-dUTP per Chain

The number of BODIPY moieties per polynucleotide chain was determined by BODIPY fluorescence measurements after enzymatic treatment with Exonuclease I (BioLabs) (Scheme S2). Exonuclease I catalyzes the removal of BODIPY-dUTP mononucleotides from ssDNA in the 3’ to 5’ direction. This fragmentation into individual BODIPY-dUTP nucleotides prevents BODIPY self-quenching, and allows for quantitative fluorescence measurements. Fluorescence measurements for the polynucleotide samples, and the Cy5 and BODIPY standard series, were carried out with a Victor™ microplate reader (Perkin Elmer Life Sciences). While the Cy5 fluorescence was excited at 650 nm and measured at 680 nm, the BODIPY fluorescence was excited at 485 nm and measured at 535 nm. The average amounts of BODIPY and Cy5 in the samples were determined by comparison to the fluorescence intensities of the standard curves. The average number of B-dUTP per polynucleotide chain was then calculated by taking the molar ratio of BODIPY to Cy5.

Scheme 10.52. Schematic illustration of Exonuclease I catalyzed hydrolysis, used in the removal of the BODIPY-dUTP nucleotides from polynucleotide chains in the 3’ to 5’ direction, yielding brightly fluorescence reaction products.
Atomic Force Microscopy (AFM) Imaging

TappingMode AFM images were acquired under ambient conditions with a MultiMode AFM (Bruker), using TappingMode silicon cantilevers \( (k_F = 40 \text{ N/m}, f_{res} = 311-357 \text{ kHz}, R_{Tip} < 10 \text{ nm}, \text{Bruker}) \). The image obtained by AFM reflects the convolution of surface topography and also the shape and size of the scanning probe tip. While tip-broadening artifacts can affect the lateral resolution, the measured height values are not significantly affected. We thus chose to measure and analyze the center height of our micelles to estimate the (collapsed) core size.

Figure 10.S2 shows the AFM TappingMode height images of star like micelles (pT1B, pT2B and pT3B). These images show the uniform size and distribution of our polynucleotide micelles over large scan areas. Figure 10.S3 shows an AFM TappingMode image of poly(dTTP) homopolynucleotides prepared and imaged under the same conditions as the BODIPY modified poly(dTTP). This image shows that the homopolynucleotides did not form micelles or aggregate.

![AFM TappingMode height images of micelles](image)

**Figure 10.S2.** AFM TappingMode height images of star like micelles (pT1B, pT2B and pT3B). These AFM images show the uniform size and distribution of our polynucleotide micelles over large scan areas.
Figure 10.S3. AFM TappingMode height image of pT2.

DLS Profiles

Figure 10.S4. $R_h$ plots from DLS measurements in 300 mM NaCl for the homopolynucleotides (pT1, pT2, pT3) and amphiphilic ssDNA samples (pT1B, pT2B and pT3B). The $R_h$ of the unmodified homopolynucleotides (corona block) increased approximately linearly from about 9 nm to 21 nm with increasing MW from 0.7 to 2.7 kb. Similarly, the micellar $R_h$ increased approximately linearly from pT1B to pT3B, corresponding to the increase in MW of the homopolynucleotides in the corona.

As expected, the $R_h$ for the micelles decreases with increasing ionic strength of the solvent, i.e., increased electrostatic screening, as shown in Table S2.
Table 10.S2. $R_h$ values of amphiphilic polynucleotide micelles at different NaCl concentrations.

<table>
<thead>
<tr>
<th>NaCl Conc. (mM)</th>
<th>0</th>
<th>50</th>
<th>100</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>pT1B $R_h$ (nm)</td>
<td>51.9 ± 0.6</td>
<td>38.0 ± 1.0</td>
<td>36.1 ± 0.5</td>
<td>32.9 ± 0.8</td>
</tr>
<tr>
<td>pT2B $R_h$ (nm)</td>
<td>93.3 ± 14.6</td>
<td>51.7 ± 0.9</td>
<td>51.4 ± 0.9</td>
<td>45.9 ± 1.5</td>
</tr>
<tr>
<td>pT3B $R_h$ (nm)</td>
<td>97.3 ± 4.0</td>
<td>60.5 ± 2.2</td>
<td>59.4 ± 1.4</td>
<td>56.6 ± 0.4</td>
</tr>
</tbody>
</table>

CMC Determination

The critical micelle concentration (CMC) for our systems is defined as the concentration above which amphiphilic polynucleotides assemble into micellar structures. Here we determine this concentration from the maximum change in the concentration-dependence of the BODIPY fluorescence intensity (Eqs. 3 and 4). This approach is reasonable because the measured BODIPY fluorescence intensity is a linear function of the concentration of all the amphiphiles involved in the micellization process.

\[
\left(\frac{d^3 \phi}{d C_T^3}\right)_{C_T=CMC} = 0, \quad (3)
\]

and

\[
\phi = A[C_s] + B[C_m], \quad (4)
\]

where $C_T$ is the total polynucleotide amphiphile concentration, $\phi$ is the weighted BODIPY intensity arising from the concentration of single chains [$C_s$] and micelles [$C_m$], and $A$ and $B$ are proportionality constants. The measured data were first fit with polynomial functions, and the CMC was then determined by setting the 3rd derivative of the polynomial equal to zero.\(^2\)
Simulations and Results

Simulation Method

Dissipative Particle Dynamics (DPD), a meso-scale simulation technique, is used to study and predict the phase behavior and properties of copolymer aggregates as well as complex, self-assembled structures such as multicomponent micelles and polymersomes.\[3\] In DPD, a number of molecular entities are coarse-grained into an element, thereafter called a DPD bead.\[3b, 4\] These DPD beads move according to Newton’s equations of motion. For a DPD bead \(i\), we have

\[
\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \quad m_i \frac{d\mathbf{v}_i}{dt} = \sum_{j \neq i} \mathbf{f}_{ij}, \quad \mathbf{f}_{ij} = (\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R) + \mathbf{F}_{ij}^S
\]

where \(m_i\), \(\mathbf{r}_i\) and \(\mathbf{v}_i\) are the mass, position, and velocity of bead \(i\), respectively. \(\mathbf{f}_{ij}\) is the interbead force on bead \(i\) by bead \(j\). \(\mathbf{F}_{ij}^C\), \(\mathbf{F}_{ij}^D\), and \(\mathbf{F}_{ij}^R\) are the conservative force, dissipative force, and the random force, respectively, acting between beads \(i\) and \(j\). They are given by

\[
\mathbf{F}_{ij}^C = \begin{cases} a_{ij} \left(1 - \frac{r_j}{r_c} \right) \hat{\mathbf{r}}_{ij}, & r_j < r_c \\ 0, & r_j \geq r_c \end{cases}, \quad \mathbf{F}_{ij}^D = -\gamma w^D(r_j) (\hat{\mathbf{r}}_{ij} \cdot \mathbf{v}_j) \hat{\mathbf{r}}_{ij}, \quad \mathbf{F}_{ij}^R = \sigma_p w^R(r_j) \theta_{ij} \Delta t \mathbf{\hat{r}}_{ij}
\]

where \(a_{ij}\) is the maximum repulsion between beads \(i\) and \(j\); \(\mathbf{r}_j = \mathbf{r}_i - \mathbf{r}_j\), \(r_j = |\mathbf{r}_j|\), \(\mathbf{v}_j = \mathbf{v}_i - \mathbf{v}_j\), and \(\hat{\mathbf{r}}_{ij} = \mathbf{r}_j / r_j\) is the unit vector directed along \(j\) to \(i\). \(\gamma\) and \(\sigma_p\) are coefficients characterizing the strengths of the dissipative and random forces. \(\Delta t\) is the iteration time step. \(\theta_{ij}(t)\) is a symmetric random variable.

Español and Warren\[5\] showed that for the fluctuation-dissipation theory to be satisfied at
some temperature $T$, the two weight functions and their prefactors must satisfy the following relationship:

$$w^D(r) = [w^R(r)]^2, \quad \gamma = \frac{\sigma^2_{\text{br}}}{2k_BT}, \quad w^D(r) = [w^R(r)]^2 = \begin{cases} (1 - r/r_c)^2, & r < r_c \\ 0, & r \geq r_c \end{cases}$$

In DPD, the properties of the system are expressed using dimensionless quantities in units of the cutoff $r_c$, the energy scale $k_BT$ and the bead mass $m_0$. As a consequence, the unit of time $\tau$ is $\tau = \sqrt{r_c^3 m_0 / k_BT}$. All simulations were carried out with the LAMMPS program package (Sandia National Laboratory).\textsuperscript{[6]}

**Simulation Set-up**

The simulation system is comprised of amphiphile chains and water molecules. In our system, the amphiphile chains are represented as a bead-spring type particle model, where adjacent beads in the chains are connected via an extra harmonic spring:

$$F_{ij}^S = Cr_{ij},$$

where the spring constant $C$ is set to 4.0 to link polymer beads together in the backbone. The choice of $C$ will not affect the qualitative behavior of the system. When compared with dsDNA, which typically is described by a worm-like chain (WLC) model, ssDNA's flexibility is much higher and solvent dependent; i.e., the persistence length of ssDNA decreases with increasing salt concentration. We thus assumed that at the salt concentration of 0.3 M NaCl used in the experiments, ssDNA can still be described as a flexible chain and we thus exclude angle and dihedral angle potentials in the simulations. Water is also modeled by coarse-grained beads, each corresponding to a group of several $\text{H}_2\text{O}$ molecules. The repulsion parameters in our system were parameterized based on the Groot and Warren\textsuperscript{[3b]} scheme, where the repulsion parameter for DPD beads of the same type is equal to $a_{ii} = 25k_BT$ to ensure mutual solubility and the bead density of the system, $\rho = 3$, corresponds to a liquid at room temperature. The negative charges of the polynucleotide chain segments were considered to be screened, reflecting the high ionic strength used in the experiments. The repulsive parameters for the hydrophilic-hydrophobic segment interactions in and between polynucleotide chains and between the hydrophobic polynucleotide segments and water molecules were set to 200 $k_BT$ to obtain aggregates in water. The interaction parameter between the hydrophilic polynucleotide segments and water molecules were set to 26 $k_BT$. To reach dilute solution conditions, only ten amphiphile chains were introduced in each simulation system.
Figure 10.S6. DPD simulation set-up.

All DPD simulations started from a random dispersion of the amphiphile chains in solution. A cubic simulation box was applied with periodic boundary condition in three directions (Figure 10.S6). DPD simulations were performed in a cubic cell of size $40 \times 40 \times 40 \, r_c^3$, where $r_c$ is the DPD unit of length. The total number of beads in the system was 192,000. To investigate the effect of polynucleotide block length on assembly behavior, we varied the hydrophilic block length, $N_A = 300, 500, 600$ and 900, while keeping the hydrophobic block length the same, $N_B = 4$. The volume fraction of amphiphiles was 0.0158, 0.0262, 0.0315, and 0.047 for polynucleotide block length of 300, 500, 600 and 900, respectively. Time evolution of the system was calculated by the Verlet algorithm, with a time step of 0.05$\tau$, where $\tau$ is the DPD unit of time. A considerably long simulation ($> 6 \times 10^6$ steps) was performed for each group to attain thermodynamic equilibrium.

Results

In this study, we examined the micellization behavior of polynucleotide amphiphiles with very long hydrophilic blocks. To the best of our knowledge, ours chains are the longest DPD models used to date. The aggregation number ($m_{sim}$) and the micellar size ($R_g$) were calculated to illustrate the effect of polynucleotide block length on micellar morphology.

Depending on the length of the hydrophilic block, two different kinetic processes of micellar assembly were observed. With relatively short hydrophilic chains, small polynucleotide
aggregates form that eventually self-assemble into micelles (Figure 10.S7 a-d). With increasing length of the hydrophilic chains, the aggregation of small micelles becomes more difficult, and a slower process of an association-dissociation equilibration of the biggest micelle was seen (Figures 10.S7e-h, Figure 10.S8). With increasing hydrophilic block lengths the aggregation number, $m_{\text{sim}}$ initially decreases from 9 to 5 and then appears to stabilize at 5 for block lengths exceeding 600 nucleotides. A crossover from the mean-field (shorter ssDNA lengths) to the scaling behavior (longer ssDNA lengths) occurs in our simulations at around 500-600 nucleotides which is also consistent with a previous experimental study (Table 10.S3).\textsuperscript{[7]}

Scaling theory predicts that the aggregation number and micellar core radius mainly depend on the size of the hydrophobic block.\textsuperscript{[8]} In contrast to scaling theories, however, the mean field theory formulated by Noolandi et al.\textsuperscript{[9]} and Nagarajan and Ganesh\textsuperscript{[10]} predicts that the hydrophilic block can have a potentially strong influence on micellization and micellar morphology. The size of the micelle and of the micellar core, and the aggregation number, all can decrease even though the length of the hydrophilic block increases. This is due to a larger loss of conformational and translational entropy for a longer chain when it becomes part of a micelle. The decrease of aggregation number and micellar size with increasing hydrophilic block length has been observed in several DPD simulations.\textsuperscript{[3d, 11]} However, in these simulations, relatively “short” block copolymers with block lengths between 2 and about 60 repeat units were considered. The decrease in micellar size with increasing hydrophilic block length has also been observed experimentally. For example, in a SANS study of PEO-PEP copolymers with a wide range of PEO block lengths and constant PEP block length, the observed decrease in aggregation number with increasing hydrophilic block length, reflected the transition from the mean-field to the scaling behavior for star-like micelles. For sufficiently long PEO blocks, the aggregation number became independent from the length of the hydrophilic block.\textsuperscript{[7]} Since in our simulations we also have very long chains, we were able to observe a similar crossover between scaling theory and mean field theory at a length of about 600 DPD beads (Table S3).

To estimate the sizes of micelles, the radius of gyration of the biggest micelles and their micellar cores were calculated according to the positions of block copolymer beads every $10^3$ steps. The radius of gyration of the biggest micelle does vary similarly, as shown in the main text. Furthermore, the micellar core radius decreases at first, but then stays approximately constant regardless of the hydrophilic block length (Figure 10.S9). Micelles with small aggregation numbers form when the hydrophilic block is long enough and the hydrophilic
block can coil back to contact with the hydrophobic core. The core size of the micelles is practically the same for all four simulation cases.

### Table 10.5.3. Aggregation number ($m_{\text{sim}}$) as a function of hydrophilic block length ($N_A$).

<table>
<thead>
<tr>
<th>$N_A$</th>
<th>300</th>
<th>500</th>
<th>600</th>
<th>900</th>
<th>Experimental estimates$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum $m_{\text{sim}}$</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>3.7-4.7</td>
</tr>
</tbody>
</table>

$^a$ Experimental estimates counted from AFM images are 3.7, 4.7 and 4.2 corresponding for pT1B, pT2B, and pT3B respectively.

**Figure 10.5.7.** Snapshots from DPD simulation trajectories with hydrophilic block lengths of (a-d) 300 and (e-h) 900 nucleotides, respectively. The chains are colored according to their aggregation number where blue represents the largest micelle, green represents an aggregation number of two, and pink reflects a single chain.
Figure 10.S8. Dynamics of micelle formation in the beginning of the simulations. (a-c) snapshots of the system taken at different time on (d). (d) Number of clusters versus time colored by the aggregation number.

Figure 10.S9. Radius of gyration of the micelles core. (a) snapshot from the simulations after $10^6$ times steps. (b) Core $R_g$ plotted as a function of hydrophilic block length for two aggregation numbers.
References