11-12-2021

Computationally Driven Design of Shape-Controlled Polymeric Nanostructures

Davindra Tulsi
University of South Florida

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Chemical Engineering Commons

Scholar Commons Citation
Tulsi, Davindra, "Computationally Driven Design of Shape-Controlled Polymeric Nanostructures" (2021).
Graduate Theses and Dissertations.
https://digitalcommons.usf.edu/etd/9247

This Dissertation is brought to you for free and open access by the Graduate School at Digital Commons @ University of South Florida. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact scholarcommons@usf.edu.
Computationally Driven Design of Shape-Controlled Polymeric Nanostructures

by

Davindra Tulsi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering
Department of Chemical, Biological, and Materials Engineering
College of Engineering
University of South Florida

Major Professor: David Simmons, Ph.D.
Norma Alcantar, Ph.D.
Clifford Henderson, Ph D.
Nathan Gallant, Ph.D.
Sameer Varma, Ph.D.

Date of Approval:
November 2, 2021

Keywords: Single-molecule nanoparticle, sequence-controlled polymers, molecular shape control, artificial vesicle, Brownian dynamics simulations

Copyright © 2021, Davindra Tulsi
Dedication

I dedicate this work to my mother and my aunt for both supporting me financially and mentally throughout this entire journey.
Acknowledgments

I would like to express my gratitude to all the people who have played a part in helping me throughout this entire journey. I would like to thank my advisor, Dr. David S. Simmons for his continuous support and mentorship. It was a pleasure working with him, and I am grateful for his patience in the beginning as I had little to no knowledge of computational techniques. I would also like to thank Dr. Clifford Henderson, Dr. Norma Alcantar, Dr. Sameer Varma, and Dr. Nathan Gallant for being a part of both my proposal and dissertation defenses. I am grateful for all the feedback they have presented to me, which ultimately strengthened the research I conducted within the five years.

I would like to thank all members of the Simmons research group for maintaining a healthy and productive working environment. I am especially grateful to past group members, Dr. Venkatesh Meenakshisundaram and Dr. Tarak Patra in mentoring me on computational methods, genetic algorithms, and machine learning within the contexts of my research. Additionally, I would like to thank my colleague, William Drayer and my friend, Jakin Delony for their continuous support and feedback.

I would like to extend my gratitude to the great friends that I have met in Ohio and Tampa during my stay in the US. I especially would like to thank my friends, Jonathan and Saksham who supported me when I needed encouragement the most. Finally, I would like to thank my family for their support and motivation throughout my graduate studies.
This work was supported in part by the National Science Foundation (NSF) Career Award under Grant No. DMR - 1849594.
# Table of Contents

List of Tables ........................................................................................................... iii

List of Figures ........................................................................................................... iv

Abstract .................................................................................................................... viii

Chapter 1: Background and Literature Review ...................................................... 1
  1.1 Shape Control in Synthetic Macromolecules .............................................. 5
  1.2 Sequence-Controlled Polymers .................................................................. 12
  1.3 Single Chain Technologies ......................................................................... 13
  1.4 Theory ........................................................................................................... 15
  1.5 Genetic Algorithms ...................................................................................... 15

Chapter 2: General Methodology ......................................................................... 20
  2.1 Brownian Dynamics .................................................................................... 20
  2.2 Measures of Shape ....................................................................................... 24
  2.3 Clustering ...................................................................................................... 25
  2.4 Neighbor Count Analysis ............................................................................ 25

Chapter 3: Design of Single Chain Nanoglobules .............................................. 27
  3.1 Determination of the Design Space ............................................................. 27
  3.2 Effect of Solvophobic Block Length on Nanoglobal Shape ....................... 38
  3.3 Effect of Interfacial Length on Nanoglobal Shape ....................................... 45
  3.4 Effect of Stiffness on Nanoglobal Shape ................................................... 45
  3.5 Comparison to Multimolecular Assembly .................................................. 49
  3.6 Elucidating a Sequence-Shape Relationship .............................................. 50
  3.7 Summary and Conclusions ......................................................................... 52

Chapter 4: Design of Aggregation-Resistant Single Chain Vesicles ................ 55
  4.1 Introduction ................................................................................................... 55
  4.2 Methodology ................................................................................................ 56
  4.3 Aggregation Behavior of ‘Bare’ Nanoglobal Shapes ................................... 57
  4.4 Design of Aggregation-Resistant Vesicles ................................................ 58
  4.5 Aggregation Behavior .................................................................................. 61
  4.6 Summary and Conclusions ......................................................................... 66

Chapter 5: Design of Motif Multiblock Copolymer Globules ............................ 68
  5.1 Introduction ................................................................................................... 68
List of Tables

Table 2.1. Lennard Jones parameters for thermodynamic routes to achieve globular states........................................................................................................................................22

Table 2.2. Lennard Jones parameters for multichain assembly (‘chopped’ 18mers)........23

Table 3.1. Genetic algorithm parameters ..................................................................................29

Table 3.2. Repeating sequences of necklace conformations with total length 10 and their respective block length distributions ..........................................................33

Table 3.3. Top 10 anisotropic sequences from the genetic algorithm.................................35

Table 3.4. Repeating motifs of form $A_\alpha B_\beta$ utilized in this work with DOP $\approx 2000$........38

Table 4.1. Lennard Jones parameters for multichain assembly of 2000 mer globules .........56

Table 4.2. Sequences of grafted single chain vesicles with vesicle motif, $A_3B_2$ ..........58

Table 4.3. Vesicles studied at low concentration with sequence, $[(A_6B_2)_xA_2B_y]_{n-1-(A_6B_2)_x}A_z$ ..........................................................60

Table 5.1. Lennard Jones parameters for thermodynamic routes for motif diblocks..........68

Table 6.1. Fitness values for various molecular shapes at different voxelization resolutions against a target cylindrical shape...............................................................84

Table 6.2. Fitness values for various molecular shapes at different voxelization resolutions against a target curved cylindrical shape..................................................84

Table A.1. Repeating sequences that exhibit cylindrical-toroidal conformations and their respective block length distributions at DOP = 500.........................................100

Table A.2. Exponential fitting parameters for vesicle with repeating motif (i) $A_6B_2$, and $[(A_6B_2)_xA_9(B_y)_{n-1-(A_6B_2)_x}A_9$ for (ii) $x = 25$, $y = 50$, and (iii) $x = 25$, $y = 75$, and (iv) $x = 10$, $y = 20$. ..................................................................................109

Table A.3. Motif diblock copolymer globules simulated with total DOP $\approx 5600$............110
List of Figures

Figure 1.1. Schematic of a genetic algorithm highlighting evolutionary operators ..................16

Figure 2.1. Thermodynamic routes, (I) and (II) employed to achieve nanoglobular shapes. ..........................................................21

Figure 2.2. Demonstration on the DBSCAN Clustering Algorithm from Scikit-Learn in determining the number of solvophobic clusters in two nanoglobular conformations, (a) string of connected beads, and a (b) toroidal conformation.......26

Figure 3.1. Number of unique sequences as a function of the length of a binary copolymer ..........................................................................................27

Figure 3.2. (a) Plot of radius of gyration of conformation for homopolymer globules against the reduced temperature of the system........................................28

Figure 3.3. Conformations of single chain block copolymer globules with sequence, (AB)N where N = 1 to 50. .................................................................30

Figure 3.4. The relationship between the relative shape anisotropy and fraction of solvophobic beads for all 10 mer repeating motifs with overall DOP on left (a) 100 and right (b) 250. ...........................................................................31

Figure 3.5. Typical globular conformations with repeating motifs (a) A2B8, (b) A4B6, (c) A3B7, (d) A3B2, (e) A1B2A2B2 and (f) A3B3A3B1 at DOP = 250..................................................32

Figure 3.6. Results from a genetic algorithm to design a repeating sequence that permits folding into an anisotropic globule. .................................................................34

Figure 3.7. Neighbor counts for repeating sequence, A3B2A2B3. Values are averaged over all 200 repeating beads for a total DOP of 2000. ........................................36

Figure 3.8. Effect of degree of polymerization on molecular shape for single copolymer chain globules with repeat units: (I) A3B2, (II) A7B3 and (III) A8B2. ..................................................37

Figure 3.9. Conformation diagram of single chain copolymer globules with repeat unit, AαBβ at DOP ≈ 2000. .............................................................................39

Figure 3.10. Quantitative analysis of sequences with repeat unit, α + β = 10: (b) dependence of aspect ratio (L1 = λ3/λ1) on solvophobic block length and (c) number of average clusters on solvophobic block length as defined using the DBSCAN clustering algorithm in Scikit-Learn. ........................................40
Figure 3.11. Effect of degree of polymerization on molecular shape for single copolymer chain globules with repeat units: (I) $A_3B_2$, (II) $A_7B_3$ and (III) $A_8B_2$.

Figure 3.12. Effect of DOP on aspect ratio for three repeating sequences: $A_3B_2$ (blue circles), $A_7B_3$ (orange squares) and $A_8B_2$ (green triangles).

Figure 3.13. Number density ($\rho$) binned from the center of the mass ($r$) for macromer, $A_3B_2$.

Figure 3.14. Dependence of vesicle (i) wall thickness and (ii) solvophilic cavity radius on solvophobic block length ($\alpha$) for solvophilic block lengths ($\beta$) ranging from 1 to 4 as indicated in the inset.

Figure 3.15. Non-normalized radial distribution function plots for sequences with repeating motifs, $A_\alpha B_\beta$, for (i) $\alpha + \beta = 6$, (ii) $\alpha + \beta = 8$, (iii) $\alpha + \beta = 10$, and (iv) $\alpha + \beta = 12$.

Figure 3.16. Conformation diagram of single chain copolymer globules with repeat unit, $A_\alpha B_\beta$ at DOP = 2000 at varying A-B attractive interaction: (i) $\epsilon = 1$, (ii) $\epsilon = 1.5$ and (iii) $\epsilon = 2$.

Figure 3.17. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 5$ at $\epsilon = 2.0$.

Figure 3.18. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 5$ at $\epsilon = 1.5$.

Figure 3.19. Comparison of conformations of multimolecular assemblies (with a drag force of 0.01 LJ units) with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$.

Figure 4.1. Aggregation behavior of single chain copolymer globules with repeating motifs: (a) $A_3B_2$, (b) $A_3B_7$ and (c) $A_5B_5$ for simulation protocols without and with a drag force.

Figure 4.2. Vesicles with sequence, (i) $B_{200}(A_3B_2)_{400}$, (ii) $B_{100}(A_3B_2)_{400}B_{100}$, (iii) $[B_{25}(A_3B_2)_{100}]_4$, [(A$_6$B$_2$)$_{40}$B$_y$]$y$-(A$_3$B$_2$)$_{40}$ with (iv) $y = 50$, (v) $y = 75$, and (vi) $y = 100$.

Figure 4.3. Simulation images, rendered in VMD, are shown in (b) for vesicle with repeating motif (i) $A_6B_2$, and [(A$_6$B$_2$)$_{x}$A$_9$(B$_3$)$_{y}$]$_{x}$-(A$_6$B$_2$)$_{y}$A$_9$ for (ii) $x = 25$, $y = 50$, (iii) $x = 25$, $y = 75$, (iv) $x = 10$, $y = 20$, (v) $x = 10$, $y = 40$, and (vi) $x = 10$, $y = 60$.

Figure 4.4. (a) Aggregation behavior of various vesicles shown in (a) via tracking the number of solvophobic clusters over time as defined using the DBSCAN clustering algorithm in Scikit-Learn.
Figure 4.5. Exponential curve fitting (green lines) to aggregation behavior for vesicle with repeating motif (i) $A_6 B_2$, and [(A$_6$B$_2$)$_x$ A$_9$ (B$_3$)$_{x-1}$- (A$_6$B$_2$)$_x$ A$_9$ for (ii) $x = 25$, $y = 50$, and (iii) $x = 25$, $y = 75$, and (iv) $x = 10$, $y = 20$. ........................................64

Figure 4.6. Aggregation behavior of vesicle with repeating motif, A$_6$B$_2$, in the absence of a drag force, via tracking the number of solvophobic clusters over time as defined using the DBSCAN clustering algorithm in Scikit-Learn.......................................65

Figure 5.1. Motif diblock copolymer, (A$_6$B$_2$)$_{350}$-b-(A$_4$B$_4$)$_{150}$, with total degree of polymerization 4000 comprised of blocks individually coding for a vesicle and a worm, respectively..................................................71

Figure 5.2. Bar chart highlighting pore duration times for thermodynamic routes I (left) and II (right). ........................................................................................................72

Figure 5.3. Motif diblock copolymer globules, (A$_6$B$_2$)$_{350}$-b-(A$_4$B$_4$)$_m$, comprised of blocks individually coding for a vesicle and a worm, respectively........................................73

Figure 5.4. Motif diblock copolymer globules, (A$_6$B$_2$)$_{350}$-b-(A$_7$B$_8$)$_m$, comprised of blocks individually coding for a vesicle and a worm, respectively........................................75

Figure 5.5. Simulation images rendered in VMD of poreated vesicles with sequences (a) (A$_6$B$_2$)$_{350}$-b-(A$_4$B$_4$)$_{175}$, and (b) (A$_6$B$_2$)$_{350}$-b-(A$_3$B$_3$)$_{133}$..................................................76

Figure 5.6. Bar plot highlights the areas for the two poreated vesicles over the two thermodynamic routes: route I (colored red) wherein collapse preceeds AB segregation and route II (blue) wherein collapse and segregation occurs simultaneously.................................................................76

Figure 5.7. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and a worm, respectively..................................................77

Figure 6.1. Visual rendering of voxel maps of the target cylindrical shape (red) against (i) sphere, (ii) cylinder, (iii) toroid, and (iv) curved cylinder query molecular (all in blue) shapes........................................................................................................85

Figure 6.2. Visual rendering of voxel maps of the target curved cylindrical shape (red) against (i) sphere, (ii) cylinder, (iii) toroid, and (iv) curved cylinder query molecular (all in blue) shapes........................................................................................................85

Figure 6.3. Neural-network biased genetic algorithm for inverse material design..................................................85

Figure A.1. Cylindrical-toroidal conformations seen for repeating sequences in the above table.................................100

Figure A.2. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, A$_{\alpha}$B$_{\beta}$ for $\alpha + \beta = 10$ at $\epsilon = 1.5$ with $6 < \alpha < 10$. .................................................101

Figure A.3. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, A$_{\alpha}$B$_{\beta}$ for $\alpha + \beta = 10$ at $\epsilon = 1.5$ with $3 < \alpha < 7$. .................................................102
Figure A.4. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $7 \leq \alpha \leq 11$..................................................................................................................103

Figure A.5. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $1 \leq \alpha \leq 6$.................................................................................................................................104

Figure A.6. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $15 \leq \alpha \leq 17$........................................................................................................................................105

Figure A.7. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $12 \leq \alpha \leq 14$........................................................................................................................................106

Figure A.8. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $7 \leq \alpha \leq 11$........................................................................................................................................107

Figure A.9. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $1 \leq \alpha \leq 6$........................................................................................................................................108

Figure A.10. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, $(A_4B_4)_{350}$........................................................................................................111

Figure A.11. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, $(A_5B_5)_{280}$........................................................................................................112

Figure A.12. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, $(A_6B_6)_{233}$........................................................................................................113

Figure A.13. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, $(A_6B_4)_{280}$........................................................................................................114
Abstract

In the realm of shape control involving synthetic soft materials, block copolymers are ubiquitously exploited due to their ability to spontaneously self-assemble into a plethora of morphologies. The growing need in areas such as drug delivery, lithography, and microelectronics to generate new nanostructures demands ever finer control of copolymer self-assembly. It is desirable to better control the shape and size of soft nanostructures, with minimal post processing techniques to assert stability.

While there is an abundance of literature on diblock copolymer assembly, few relationships exist for sequence-controlled copolymers. It is well known that biological macromolecules are highly sequence specific and exploit secondary interactions to direct the assembly into a hierarchy of structures with controlled shape and size. Single chain synthetic sequence-controlled copolymers that self-assemble into structures with well-defined shape, size, and stability could extend the design space of accessible shapes formed via traditional diblock copolymer self-assembly, and potentially offset the time and cost required for post-modification processes to assert stability. However, employing an in-depth study through trial and error is not practical; thus, emphasizing the need for the development of an efficient structure-searching strategy.

Inspired by nature’s efficiency in self-assembly and shape recognition, the combined efforts of evolutionary algorithms, computer vision and Brownian dynamics simulations were used
as a meta-heuristic approach to search and design sequence-controlled polymers that assemble into targeted shapes. This work is grouped into four main areas:

1. Design of single chain nanoglobules: To provide a fundamental understanding on the design of single chain nanoglobules with targeted shapes
2. Design of aggregation resistant globules: To design single chain vesicles that are resistant to aggregation at low concentration
3. Design of hierarchial nanoglobules: To utilize single chain nanoglobules as building blocks for the design of hierarchical nanostructures
4. Design and implementation of a structure-searching strategy

In the first part of this work, the effects of the following length scales on nanoparticular shape were investigated based on a bead-spring polymer model:

1. A/S Interfacial length (ε)*
2. Kuhn length (b)

Results yield a conformational diagram mapping molecular sequence to single molecule nanoparticular shapes ranging from vesicles to sponges to necklaces. Here, chain sequence can offer fine control of nanoparticulate dimensions, for example enabling control of vesicle cavity size. Within this model, complex shape control is dictated by a large molecular weight and operating in the limit of strong A-B segregation (high χ_{AB}).

Further, single chain globules are designed that are resistant to aggregation at low concentration. Results indicate that the attachment of solvophilic loops to single chain vesicles in
a sequence-specific manner permits steric repulsion that prevents aggregation; offering kinetic stabilization within the timescales probed.

The single chain nanoglobules designed in the first part of the work provide an ‘alphabet’ of nanoparticular shapes for hierarchical shape design in the penultimate part of this work. The combination of multiple sequence ‘motifs’ consisting of vesicle and worm motifs demonstrated the design of poreated and tubular vesicles, pointing the way towards designer artificial enzymes and tubules. It was further shown that the pore size of poreated vesicles can be directly tuned by controlling the sequence of the worm-coding block.

In the final part of the work, a shape-matching algorithm is developed to search for targeted shapes based on a 3D voxelization scheme. The shape-matching algorithm was tested with various model shapes and retrofitted for use in a genetic algorithm for inverse material design.
Chapter 1: Background and Literature Review

Shape is a key parameter in predicting the structural and dynamical properties of materials. It affects how molecules arrange, connect, and assemble into larger structures\(^1\). Additionally, the hydrodynamic behavior of materials depends on their shape\(^2\). Star-shaped polymers, for instance, are used to tune the rheological properties of lubricating oils\(^3\).

Realization of controlled assembly of synthetic molecules comparable to that achieved in biological macromolecules has been a long-term goal of polymer science\(^4\). Here, the literature review focuses on the importance of macromolecular shape control, conventional techniques using multimolecular assembly and secondary interactions. Motivated by shape control in biological macromolecules modulated by sequence control of single chain macromolecules, literature on sequence-controlled polymers, single chain technologies, followed by the theoretical and practical design challenges of sequence-controlled polymers are presented. Finally, a review of the genetic algorithm and its application as a structure-searching strategy is presented.

1.1 Shape Control in Synthetic Macromolecules

From the artificial standpoint, efforts in designing nanoparticular shapes have predominantly relied upon a balance between long-ranged interactions\(^5,6\) and short-ranged solvation forces\(^7\) or have employed multi-molecular assembly\(^8-11\).

Driven by the incompatibility of blocks with different solubility, block copolymers micro-segregate to form a plethora of morphologies. Pioneering work by Helfand\(^12\) and Leibler\(^13\) led to the development of phase diagrams to predict various morphologies\(^14\) based on the
thermodynamic segregation parameter ($\chi N$), and composition ($f$). In the mean field treatment of a homogeneous linear block copolymer in a melt with incompatible blocks, A and B and respective degrees of polymerization, $N_A$ and $N_B$, Leibler predicted various morphologies such as lamellar, body-centered-cubic, hexagonal and inverted hexagonal.\textsuperscript{15} More complex morphologies in bulk can be realized by manipulation of polymer architecture and secondary interactions such as electrostatics.\textsuperscript{16}

The growing need in areas for complex shape control such as in drug delivery devices with targeted circulation times, drug loading/release times has motivated the study and design of block copolymer assemblies in aqueous solution. In a general sense, the block copolymer consists of solvophobic and solvophilic segments. Linear block copolymers have been shown to self-assemble into micellar shapes such as spheres, cylinders and bilayers\textsuperscript{8,17}.

In solution, manipulating processing conditions can yield non-trivial shapes. Wang and coworkers utilized a gas-liquid microfluidic reactor to produce flow-directed micellar morphologies of polystyrene-block-poly(acrylic acid) copolymers\textsuperscript{18}. Such shapes include Y-junctions, vesicles, and networks, in contrast to spherical shapes that would assemble in bulk solution.

Some non-trivial shapes can be observed via the manipulation of molecular weight. Polymerization induced self-assembly (PISA) – where the molecular weight of one block is changed, can be utilized to yield branched worms, bilayer octopi and jellyfish-like shapes.\textsuperscript{19} Jain and Bates showed that above a critical molecular weight, poly(1,2-butadiene-$b$-ethylene oxide) (PB-PEO) diblock copolymers can form Y-junctions and connect into three-dimensional networks, in contrast to vesicle and cylindrical micelles at lower molecular weights.\textsuperscript{20}
Furthermore, manipulating the architecture of individual blocks can utilized to guide the assembly into distinct morphologies. Recent work has demonstrated this in the context of coil-brush block copolymer assemblies in solution.\textsuperscript{21}

In-silico studies demonstrate that the assembly of (amphiphilic) block copolymers in solution yield vastly different phase diagrams than seen in the melt. For instance, Vasilevskaya and coworkers utilized a hydrophobic/amphiphilic (HA) model to probe the coil-to-globule transition in a selective solvent. Dissipative particle dynamics (DPD) was used to generate a phase diagram for hydrophobic-amphiphilic block copolymers.\textsuperscript{22} Here, lamellar structures are predicted to form at lower composition, in contrast to being symmetric around the critical point. These simulations have identified two non-trivial morphologies: the double diamond and hexagonal perforated lamellae (HPL).

Moving towards ternary copolymers opens opportunities towards developing new shapes and hierarchical nanostructures. With a hydrophilic block and two incompatible hydrophobic blocks, linear ternary block copolymers can self-assemble in selective solvents to yield multicompartiment micelles. For ternary copolymers with unfavorable A-C interactions, the C-blocks could be sequestered in the interior of the micelle, yielding a concentric structure. On the other hand, with more favorable A-C interactions, ternary copolymers can assemble into raspberry-like spheres with patchy surfaces.\textsuperscript{19} Moughton and coworkers\textsuperscript{23} showed that non-concentric multicompartiment micelles can be observed by changing from linear to star-like ternary block copolymers of PEO-PEO-PFPPO. These shapes include hamburger, segmented wormlike micelles and nanostructured vesicles. Mueller, Zhulina, and coworkers showed that a step-wise assembly process - involving pre-assembly in a poor solvent for B blocks followed by immersion in a poor solvent for A blocks in linear ternary block copolymers, could yield monodisperse football, clover}
and hamburger-like micelles. Cui and coworkers manipulated the kinetics of assembly of linear poly (acrylic acid)-block-poly(methyl acrylate)-block-polystyrene (PAA-b-PMA-b-PS) triblock co-polymer in aqueous solution to produce a variety of complex nanostructures such as multicompartment cylinders.

In practice, the stability of diblock copolymer assembled structures is very sensitive to processing conditions and can ultimately lead to premature disintegration in solution. To overcome this, various post-processing techniques are employed to stabilize desired shapes. One popular method involves crosslinking – either in the hydrophobic core or hydrophilic shell. Wooley, Talelli and coworkers showed that crosslinked polymeric micelles can yield biomimetic structures, where crosslinking of the hydrophobic core can yield liposomes – stable spherical shapes with a lipid bilayer shell, while shell-crosslinking can yield hollow nanoscale cage-like structures. The production of nanocages involves the initial assembly of block copolymers in aqueous solution, followed by covalent cross-linking of the hydrophobic shell. Once the shell is cross-linked, the backbone bonds of the hydrophobic core are cleaved, yielding a hollow polymer shell or nanocage. In addition to interesting properties such as high stability and greater control on permeability, cross-linked polymeric micelles can serve as a bridge towards developing more sophisticated biomimetic shapes. For instance, nanocages can be the basis for biomimetic construction of viral capsids, while core-crosslinked structures consisting of a glassy core and a charged shell can yield model structures of histones – proteins that are responsible for wrapping and ordering DNA into larger structures called nucleosomes.

In contrast to linear copolymer micelles of amphiphilic polymers, unimolecular micelles do not rely on self-assembly and are structurally stable in a wide range of environment conditions. Unimolecular micelles are defined as a class of single-molecule micelles with a distinct core and
They can be produced using dendrimers, hyperbranched, star, brush-like and amphiphilic cyclic copolymers. Since unimolecular micelles do not rely on assembly, their shapes are controlled by manipulating polymer architecture, for example, the number of arms in star polymers. For instance, Fan and coworkers synthesized a Y-shape armed unimolecular micelle from an amphiphilic star-like copolymer using a disulfide linkage at the junction points.6

Overall, the ability to control the domain spacing and functionality over a variety of length scales makes block copolymers attractive candidates for a variety of applications. In lithography, self-assembly of block copolymers can be directed onto a substrate to generate templates that exhibit periodicity at relatively small scales27. This is of growing demand due to the surge in the development of smaller electronic devices for computers.

Extending the design space of accessible shapes would enable the design of high functional materials for applications in areas such as drug delivery, catalysis and sequestration. For instance, vesicles with controlled interior and exterior residues may serve as efficient microscale reactors28; drug delivery devices can be tuned to control loading capacity and blood circulation times, while multicompartent micelles may allow preferential size sorting in sequestration/filtration applications. Moreover, the discovery of in-equilibrium shapes with advanced structural complexity would potentially offset the time and cost of post-modification processes to assert stability.

1.2 Sequence-Controlled Polymers

In contrast to multimolecular assemblies employed in synthetic macromolecular science, biological macromolecules utilize a combination of sequence control and directional interactions in single molecules to permit access to a hierarchy of nanostructures with controlled shape and
size. In a recent review on the field of sequence-controlled polymers, Lutz highlighted that the folding of single sequence-controlled polymer chains can allow access to nanostructures that are found in nature. Moreover, manipulation of sequence in single-molecule assembly of *bio-inspired macromolecules* could enable transformational new synthetic molecules for use in applications such as advanced catalysis, molecular sequestration and drug delivery.

What level of sequence control is necessary to design new materials with targeted structure and properties? The answer to this question is non-trivial at best, and first warrants a formal definition of sequence control in the context of macromolecular science. There is a large body of nomenclature surrounding the landscape of sequence-controlled polymers. To avoid ambiguity in this work, a sequence-controlled polymer is defined as a polymer wherein the monomeric sequence is controlled to a given level. This level can have many meanings, but one obvious scale is the overall length of the polymer chain. *Full sequence control* would refer to a controlled/specifed monomeric sequence along the entire chain, whereas *no sequence control* would refer to a random monomeric distribution. Perry and Sing define a sequence scale, ξ, with some analogs to the chain length scale. Within this nomenclature, a strictly binary alternating copolymer, that is, with repeating sequence, AB has a sequence scale, ξ ~ O(1), compared to ξ ~ O(N) for a binary block copolymer, where N is the degree of polymerization. Within these extremes, *sequence control* of a specified number (greater than 2 in the case of a strictly alternating sequence) of monomers can be repeated along the sequence, that is, a sequence AB that is repeated N times. In this context, such polymers have been coined as multiblock copolymers in the case where α = β, and generally, periodic copolymers. More recently, DeStefano and coworkers define precision within this context involving measures of dispersity involving molecular weight, composition, sequence and conformation.
Over the past decade, there has been significant advances in methods to synthesize sequence-controlled polymers. Lutz and coworkers (2016) categorized synthetic methods into three main areas based on the degree of sequence control/dispersity: step-growth polymerization, chain growth polymerization and multistep-growth synthetic techniques. In step-growth polymerization techniques, monomers with multiple reactive functionalities (AB) or multiple monomers (AA and BB) can be polymerized to produce periodic copolymers as described above. Compared to the other two categories, this technique usually yields in the lowest level of sequence control and molecular weight distribution. However, it should be noted that this method can yield periodic copolymers with large molecular weights via the polymerization of telechelic oligomers, that is, small polymeric species that have reactive sites at the ends of the chains. Techniques in this area have involved ring-opening metathesis to produce precise polyethylene and ionomers, step-growth click chemistry reactions, and utilization of multifunctional reactive oligomers to produce poly(lactic-co-glycolic acid) (PLGA) for use in hydrolytic degradation applications in drug delivery and release. Chain growth polymerization methods are further sub-divided into either living/controlled or uncontrolled techniques. In uncontrolled techniques, the main reactive species generally involves a growing chain with a reactive site and proceed via a ‘one pot’ or batch synthetic process. These techniques yield copolymers with a smaller sequence dispersity (greater sequence control) than step-growth polymerization. The main steps include initiation of activated species, propagation which serves to grow the chain and termination. Precision is limited by the reaction kinetics of the monomeric species, in specific, the reactivity ratios, as well as the initiation and propagation kinetics of the initial reacting chain species. Specifically, the propagation reaction kinetics can be written as follows:

\[
-M_1^* + M_1 \overset{k_{11}}{\rightarrow} -M_1M_1^*
\] (1)
\[-M_1 \cdot M_2 \xrightleftharpoons{\kappa_{12}} M_1 M_2 \cdot\]  
\[-M_2 \cdot M_2 \xrightleftharpoons{\kappa_{22}} M_2 M_2 \cdot\]  
\[-M_2 \cdot M_1 \xrightleftharpoons{\kappa_{21}} M_2 M_1 \cdot\]

where $M_1$ and $M_2$ refer to the monomer species 1 and 2 respectively; $-M_1 \cdot$ and $-M_2 \cdot$ refer to the monomer active species, and $-M_i M_j \cdot$ refers to the growing chain radical consisting of monomers $M_i$ and $M_j$. Here the active or propagating species ($\cdot$) can be based on free radical, charged (cationic or anionic) or a ring-opening mechanism.

The reactivity ratios can then be defined based on the rate constants:

\[r_1 = \frac{k_{11}}{k_{12}}\]  
\[r_2 = \frac{k_{22}}{k_{21}}\]

where, a large value of $r_i$ dictates that monomer $i$ is favored to react with itself. Several limiting cases exist. For instance, if $r_1 \approx r_2 \gg 1$, block copolymers are favored, while $r_1 \approx r_2 \approx 1$ and $r_1 \approx r_2 \approx 0$ tends to yield random and alternating copolymers respectively. Thus, the level of sequence control permitted via uncontrolled polymerization is primarily limited by the reaction kinetics of the underlying monomer species and cannot generally yield more complex sequence control than alternating and blocky sequences. In contrast, living polymerization techniques have been developed to produce a higher degree of sequence control. Here, chain termination is absent, and in the ideal case, the rate of initiation is much faster than the rate of propagation. This allows the yield of a very narrow molecular weight distribution, specifically, a Poisson distribution. Similar to uncontrolled polymerization, living polymerization can proceed via free radical such as atom transfer radical polymerization (ATRP) and reversible-addition-fragmentation chain-transfer.
polymerization (RAFT), anionic, cationic and ring-opening methods. These methods have led to the synthesis of a variety of periodic and multiblock copolymers\textsuperscript{52–55}. In multi-step-growth methods, the highest degree of sequence control or precision is permissible. Here, as the name suggests, sequence-controlled polymers are prepared via sequential or stepwise monomer addition.\textsuperscript{29} Monomers are usually bifunctional, that is, containing reactive sites from both monomeric units (A and B) for a binary copolymer. In practice, to avoid unwanted (side or chain-growth related) reactions, the multifunctional species consisting of a ‘protecting group’ attached to each reactive end. Monomer addition must therefore be preceded by a protection-deprotection reaction to mediate almost perfect sequence control. This iterative stepwise addition process can occur both in liquid and solid substrate support systems. It should be noted that purification and separation is necessary after each sequential monomer attachment in the liquid phase system. Pioneered by Merrifield\textsuperscript{56}, solid-phase synthesis proceeds with sequential monomer addition on a support of a solid media such as (originally) a crosslinked polymer, and later other solid systems such as solid nanoparticles. This allowed the growing polymer to be subsequently filtered from the reactants, without any need for batchwise filtration processes. Various biomimetic oligomers have been prepared using these techniques including peptidomimetics, foldamers, and oligonucleotides\textsuperscript{57,58}. Moving beyond oligomers, Al Ouahabi and coworkers (2015)\textsuperscript{59} demonstrated the synthesis of a 104 mer sequence-controlled polymer, polyphosphate, onto a glass pore support. The main drawback of this iterative step-growth approach is the need for high yields from individual monomeric reactions; thus, limiting the scope of possible monomers that can be utilized.

More recently, DeStefano and coworkers (2021)\textsuperscript{33} alternatively define two main synthetic categories to design sequence-controlled polymers based on either, (i) designing polymers with
defined sequence and lack of molecular weight control, or (ii) polymers with both defined sequence and molecular weight control. The first category refers to precise macromonomer methods – involving a combination of step-growth and chain-growth polymerization methods as categorized by Lutz and coworkers\textsuperscript{42} to produce mostly periodic copolymers, while the latter refers to iterative exponential growth or iterative step-growth approaches.

Overall, while there have been numerous advances in synthetic techniques, scalable production of artificial sequence-controlled polymers is still a work in progress. Iterative growth approaches on solid support systems still need non-trivial purification operations. Furthermore, the overall degree of sequence control is dependent on the kinetics of individual monomer-monomer interactions.

What is the current landscape of the use of sequence-controlled polymers for the design of shape-controlled nanostructures? Motivated by the protein folding problem, Ken Dill identified that differences in hydrophobicity (H) and polarity (P) play a significant role in dictating secondary structure of proteins as well as their stability (inability to precipitate) in solution. Using this HP model, Khokhlov and Khalatur performed molecular simulations on the self-assembly of single chain amphiphilic copolymer globules and showed that sequence can be tuned to form ‘protein-like’ globules that are resistant to aggregation/precipitation in solution\textsuperscript{7,34}. Here, protein-like sequences are designed by ‘coloring’ segments as solvophilic that preferentially move towards the outer surface of a homopolymer globule in a bad solvent. Experiments by the Segalman group\textsuperscript{28} showed that such ‘protein-like’ copolymers in the context of polypeptoids in dilute aqueous solution demonstrated a sharper coil-to-globule transition compared to regular alternating sequences. Vasilevskaya and coworkers extended the HP model to incorporate amphiphilic monomers, motivated by the fact that amino acids are amphiphilic by nature\textsuperscript{39,60}. In contrast to
lipids or diblock copolymer assembly, the connectivity between amphiphilic units in these amphiphilic homopolymers permits self-assembly at low polymer concentration. Through tuning the solvent quality and degree of polymerization, the Vasilevskaya group was able to produce a variety of structures including cylinders, necklaces, and spherical globules. Of particular importance, the group was able to design single chain vesicles and multi-layered vesicles termed as ‘onions’.

Another area that utilizes periodic sequence control is DNA nanotechnology. In DNA nanotechnology, the specific combination rules of constituent organic bases allow for a rational design of DNA strands that assemble into a target structure. As such, DNA nanotechnology relies on two important characteristics of DNA: sequence control and complementary base pairing. Since its emergence in the 1980s, advances in DNA nanotechnology have produced a variety of nanostructures with interesting shapes. Some shapes developed include polyhedra such as tetrahedra, dodecahedra, Bucky balls, ninja stars, tetrakis cubes, triakis tetrahedra, and Pentakis dodecahedra, complex shapes including various numbers, letters, and other symbols, and a variety of hierarchical superstructures.

DNA nanotechnology can be broadly grouped into two main areas: structural DNA nanotechnology, and dynamic DNA nanotechnology. In structural DNA nanotechnology, one assembly method involves the use of DNA tiles—two DNA double helices that are joined together by double strand exchanges, as building blocks for the development of nanostructures. For instance, rectangular and Y-shaped tiles can assemble via complementary base interactions at their ends to produce two-dimensional and three-dimensional lattices, as well as three-dimensional periodic structures as described above. The development of irregular two-dimensional and three-dimensional structures has been possible due to the DNA origami method. Here, a long viral DNA
strand serves as a scaffold, which via short “staple strands” is directed to fold into a target structure.\textsuperscript{17} The use of lattice origami leads to irregular two-dimensional shapes such as letters as described above, while wireframe origami – by stacking multiple layers to form a honeycomb or cubic scaffold- can be used to develop hollow three-dimensional structures.\textsuperscript{11,18} The class of dynamic nanotechnology refers to the design of nanostructures that change conformation upon some stimulus. This has led to the development of nanorobots, and drug delivery vehicles that change conformation to release a drug.\textsuperscript{11,18}

The complementary nature of DNA base pairing can also be exploited to form a variety of hierarchical structures based on colloidal particles. DNA strands can be attached to the surface as “patches”. These patchy particles can be directed to assemble into a variety of shapes. For instance, gold nanoparticles functionalized with DNA patches have been shown to assemble into satellite-like structures and tetravalent-like clusters- which bear stark resemblance to the molecular shapes of methane and ethylene.\textsuperscript{19} By varying the shape of the colloids, number, and size of patches on surfaces, colloidal particles can be programmed to self-assemble into hierarchical nanostructures and networks with symmetries such as diamond-like, disordered face-centered cubic, and orientationally ordered face-centered cubic (ofcc).\textsuperscript{19,20}

Overall, DNA nanotechnology relies on two important characteristics of DNA: sequence control and complementary base pairing. One of the drawbacks\textsuperscript{19}, however, is that \textit{shape complexity is limited by the DNA strand length}.

\subsection*{1.3 Single Chain Technologies}

A key next step in bio-inspired shape control is the assembly of targeted shape nanoobjects out of single macromolecules. The discovery of in-equilibrium shapes with advanced structural complexity would potentially offset the time and cost of post-modification processes to assert
stability as in traditional block copolymer assembly. Of particular long-term interest would be access to highly complex nanoshapes such as multi-compartment objects or multiple active sites; enabling targeted placement of moieties (such as enzymatically active sites) within a larger assembly.65

In recent years, ‘single chain technologies’ have garnered attention in developing biomimetic soft nanomaterials65. Here, inspired by nature, single polymer chains can be programmed to act as independent discrete objects or interact via bottom-up assembly to produce hierarchical structures with structural complexity. When compared to their multi-chain counterparts, single chain technologies offer smaller size, hydrodynamic volume and lower intrinsic viscosity while allowing greater control to develop multiple compartments or active sites. For instance, Terashima and coworkers reported the folding of a single terpolymer into a supramolecular structure with well-defined compartments and demonstrated catalytic hydrogenation in water66. Previous works have demonstrated that ring-like assemblies from cyclic block copolymers67,68 and covalently cross-linked polymer assemblies69 confer longer blood circulation times, decreased acid-catalyzed degradation and tunability of drug loading capacity.

1.4 Theory

The assembly of single chain synthetic sequence-controlled polymers could enable the discovery of new shapes; opening the door towards developing hierarchical nanostructures competitive with nature. Future advances in this field of sequence-controlled polymers are hindered by the fact that the relationship between sequence and molecular shape is poorly understood.
In the realm of block copolymer assembly, prediction of self-assembled shapes in selective solvents is difficult in practice. In small molecule surfactants involving a hydrophobic tail and a polar head, the interfacial free energy per molecule is defined as:

\[ \mu_N = 2\gamma a_o + \frac{\gamma}{a_o} \cdot (a - a_o)^2 \]  \hspace{1cm} (7)

where \( \gamma, a, a_o \) represent the hydrophobic-solvent interfacial tension, polar head group area and equilibrium head group area. The interplay between the hydrophobic attraction that seeks to decrease the interface and head group repulsion sets the curvature of the interface and ultimately the surfactant shape. Israelachvili showed that the transition in shapes including spheres and cylinders can be described by the packing parameter,

\[ p = \frac{v}{a_o l_c} \]  \hspace{1cm} (8)

where \( v, a_o, l_c \) represent the volume of the hydrophobic tail, equilibrium/optimum head group area and a maximum/extended length of the hydrophobic tail. For small hydrocarbon chains, no single dimension (example radius of the hydrophobic core) can exceed \( l_c \). Thus, through geometric arguments, shapes of surfactant aggregates can be described by the packing parameter. More recently, Nagarajan developed a theoretical model borrowing aspects of the surfactant assembly theory to describe the free energy per molecule of a block copolymer micelle,

\[ (\Delta\mu_g^o) = (\Delta\mu_g^o)_{A,Tr} + (\Delta\mu_g^o)_{A,def} + (\Delta\mu_g^o)_{int} + (\Delta\mu_g^o)_{B,dil} + (\Delta\mu_g^o)_{B,def} \]  \hspace{1cm} (9)

where,

1. \((\Delta\mu_g^o)_{A,Tr}\) represents the transfer free energy of hydrophobic segments from solvent to the hydrophobic core
2. \((\Delta\mu_g^o)_{A,def}\) represents the elastic deformation energy of hydrophobic segments
3. $(\Delta \mu^g)_{int}$ represents the free energy cost to create an unfavorable hydrophobic-polar interface

4. $(\Delta \mu^g)_{B, dil}$ represents the osmotic dilution of polar segments in solvent

5. $(\Delta \mu^g)_{B, def}$ represents the elastic deformation of polar segments

Here, the last two terms represent the effective ‘head’ group repulsion in the surfactant free energy model. The interplay of this head group repulsion along with the interfacial tension and elastic deformation of hydrophobic segments sets the curvature of the interface and ultimately the shape of the micelle. In regards to predicting shape, Nagarajan notes that there is no analogous geometric packing restriction can predict shape\textsuperscript{72}. There is no length scale to define a maximum extended length ($l_c$) in polymer chain conformations. Furthermore, there is no analogous free energy model to describe/account for sequence-controlled polymers.

To elicit a theoretical understanding, Khokhlov and coworkers developed a free-energy model to elucidate what factors influence the shape of protein-like copolymer globules, while noting the difficulties in developing a sequence-structure relationship\textsuperscript{73}.

While this work did not probe shape effects, combined with the literature on BCP assembly (in which coarse BCP sequence at the level of block size mediates assembly in the presence of only weak interactions), this suggests an opportunity for sequence-mediated shape control in the absence of explicit long-ranged interactions such as electrostatics.

1.5 Genetic Algorithms

The ability of experimentalists and computational researchers to collaborate and develop complex phase diagrams based on sequence control, additional monomer units, and directional
interactions could significantly extend the design space of accessible shapes achieved through traditional binary block copolymer assembly.

![Figure 1.1. Schematic of a genetic algorithm highlighting evolutionary operators](image)

However, this is a non-trivial process. Entire phase diagrams need to be mapped out for a given set of environmental to predict self-assembled morphologies. According to Lodge (2003), this begs the need for the implementation of efficient “structure searching strategies” rather than “brute force exploration” to map out the entire phase diagram.74

Evolutionary algorithms have garnered a lot of interest in solving inverse design problems involving soft matter. Previous work has involved the design of sequence-specific compatibilizers35, optimization of geometry of nanoparticles75, and the search for new crystalline structures from the self-assembly of DNA-grafted particles.76 Inspired by the Darwinist theory of
evolution, evolutionary algorithms simulate a “survival of the fitness” approach to guide the search for an optimal solution.

A schematic is depicted in Figure 1.1. While there are many different types of evolutionary algorithms, the common idea involves:

1. Mapping: The sample space of possible solutions is translated into a form that mimics the genetic coding in biology. In a genetic algorithm, the sample points are represented as strings of a finite length based on numbers or letters. For example, a binary copolymer can be mapped via a one-to-one representation into a string of 0’s and 1’s. For instance, AAAAAABB BBB can be represented as 000011111.

2. Initialization: A population of individuals (possible solutions) provide a starting point for the algorithm. This population constitutes the first generation of the algorithm. This population can be randomly generated or selected based on their proximities to the predicted optimal candidate.

3. Fitness evaluation: Every individual in the population has a certain rank or fitness based on an objective function. These values are continuously evaluated throughout every generation of the algorithm. During the first few generations, an individual with a high fitness may cause a bias in the search space. However, as the algorithm gets closer to the optimal candidate, there is a minimal variation in the fitness values in the population. To create a constant selection pressure throughout the algorithm, several scaling relationships may be employed to magnify small variations in the fitness to drive the optimization.

4. Evolution: Using the initial population, a new generation of individuals are created via selection, crossover, and mutation operators. Individuals that have a higher “fitness”, that is, a function value closer to the optimum are stochastically selected to serve as parents.
This simulates the evolutionary notion of “survival of the fittest.” Through crossover, these parents mate, and combine portions of their “genetic” material to produce a child. To maintain some diversity in the sample space, these children may be mutated, that is, portions of their genetic material may be changed based on probabilistic rules. The use of these operators is usually problem specific; some algorithms utilize selection and mutation, while others may elicit all three operators in a step-by-step manner.

5. Termination: The algorithm terminates based on a stopping criterion. This can be based on time constraints (number of evolutions or generations), or some statistics that determines if the optimal solution has been located to some acceptable level.

Artificial intelligence (AI), as the name suggests, is employed by machines to mimic aspects of human intelligence. Machine learning refers to a subset of artificial intelligence (AI), and involves the construction of algorithms that can learn and make predictions from data to solve problems of interest.79 As such, it prescribes the ability of a computer to learn without being explicitly programmed to do so. The efficiency of the biological neural network to continuously learn and access information has spawned the development of artificial neural networks (ANN) that learn by example.79

The basic building block of a biological neural network is the neuron. It receives information from branches (dendrites) originating from other neurons that protrude from its central structure. Based on an all-or nothing principle, the neuron will “fire” a signal along the axon, which in turn, carries the information until it terminates its signal to another neuron or cell.80 In a similar light, the artificial neuron can accept inputs that hold different weights. The inputs are all summed up, and based on an activation function (all-or-nothing), the output is created, which can be passed onto various neurons.
The ability of artificial neural networks to find complex relationships between multiple variables makes them attractive candidates for solving inverse design problems. In design problems where the fitness landscape may be complex with a multitude of local minima/maxima, finding the right network topology, however, is non-trivial. This was identified in a study by Tarak and coworkers. Their work involved the implementation of a neural network biased genetic algorithm (NBGA) to solve various optimization problems. One such problem involved finding a sequence-specific compatibilizer that minimizes the interfacial energy of a blend of immiscible homopolymers. In their system, the artificial neural network was a multi-layered network with each monomer in the copolymer sequence acting as a neuron in the input layer. The ANN was continuously trained by building on the fitness evaluations from molecular dynamics simulations. The ANN did not significantly accelerate the convergence of the genetic algorithm to the optimal solution. A more efficient mapping of the trained data could increase the robustness of the artificial neural networks. This can further facilitate a more efficient implementation of machine learning to constrain the design space in optimization problems involving the inverse design of soft materials.
Chapter 2: General Methodology

2.1 Brownian Dynamics

Brownian dynamics – a special case of Langevin dynamics in the absence of an average acceleration, models the stochastic movement of molecules within a fluid. To simulate this, the LAMMPS molecular dynamics simulation package is utilized with the fix langevin and fix nve commands. All simulations are employed in Lennard Jones (LJ) reduced units; a system temperature of \( T = 1.0 \). A Verlet integrator is employed with a timestep of 0.005 LJ time units.

Polymer chains simulated comprised of an attractive bead-spring polymer model. Non-bonded interactions are modeled via the Lennard-Jones (LJ) 12-6 potential:

\[
E(r) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^{6} \right]
\]  

(10)

where \( \varepsilon_{ij} \) and \( \sigma_{ij} \) correspond to energy and range parameters respectively. The interaction is truncated and shifted at a cutoff distance of \( r_{\text{cut}} = 2.5\sigma \) for attractive interactions or \( r_{\text{cut}} = 2^{1/6}\sigma \) for purely repulsive interactions. Here, \( \sigma_{ij} = 1 \) in all cases. Bonded interactions are modeled via the finitely extensible nonlinear elastic (FENE) potential,

\[
E(r) = -0.5KR_0^2 \ln \ln \left( 1 - \left( \frac{r}{R_0} \right)^2 \right) + 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right] + \varepsilon
\]  

(11)

All bonds (AA, AB, BB) have the same parameters: \( K = 30.00, R_0 = 1.0, \varepsilon = 1.00, \sigma = 1.00 \). With an implicit solvent, collapse of solvophobic beads driven by an effective self-attraction of magnitude \( \varepsilon_{AA} = 2 \) sets the solvent quality for solvophobic units. Expansion of solvophilic beads
is driven by employing purely repulsive solvophilic self-interactions and A-B cross interactions ($\varepsilon_{AB} = \varepsilon_{BB} = 1$, but with a cutoff distance of $2^{1/6} \sigma$).

Equilibrium chain conformations for single-molecule assembly are obtained via two entirely independent thermodynamic pathways to ensure equilibrium: one in which chain collapse is followed by A-B segregation (the “globule first” route); and one in which chain collapse occurs simultaneously with A-B segregation (the “simultaneous” route). Agreement of the results of these two distinct routes is taken as evidence that the system has reached equilibrium.

Specifically, in route I, the polymer starts in an expanded coil configuration; since $\varepsilon_{ij} = 0.2$ initially for all interactions, this is a homopolymer expanded coil. Thereafter, $\varepsilon_{ij}$ is ramped from 0.2 to 2.0 to induce collapse into a homopolymer globule. $\varepsilon_{ij}$ is always changed using a linear ramp function. The identity of solvophilic (B) segments are ‘switched on’ as $\varepsilon_{AB}$ and $\varepsilon_{BB}$ are decreased from 2.0 to 0.2; finally, strong repulsive interactions are turned on by changing $\varepsilon_{AB}$ and $\varepsilon_{BB}$ to

![Figure 2.1](image)

Figure 2.1. Thermodynamic routes, (I) and (II) employed to achieve nanoglobular shapes. Beads A and B are colored orange and blue respectively.
1.0, and the cutoff distance from $2.5\sigma$ to $2^{1/6}\sigma$. In route II, the system initially exists in a copolymer expanded coil configuration, and $\varepsilon_{AA}$ is subsequently ramped from 0.2 to 2.0 to induce collapse into the globular state. In both cases, a final production period of 10,000 $\tau$ is employed for visual and quantitative analysis. VMD$^{86}$ is utilized for visualization and image rendering.

Table 2.1 highlights the parameters for all pairs and simulation timescales for multichain assembly of 111 copolymers each repeating motifs of length 18: corresponding to the matrix of $\alpha + \beta = 18$ and a DOP = 1998. This route is equivalent to the “simultaneous” route, with an additional step (prior to the final globular equilibration) to enforce a small drag force (0.01 and 0.1 LJ units) to drag all molecules to their common center of mass. In the case of multimolecular assembly without the drag force, the same annealing period is employed.

Table 2.1. Lennard Jones parameters for thermodynamic routes to achieve globular states.

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
<th>$\varepsilon_{AA}$</th>
<th>$\varepsilon_{AB}$</th>
<th>$\varepsilon_{BB}$</th>
<th>$r_{\text{cut}}$ (AA)</th>
<th>$r_{\text{cut}}$ (AB)</th>
<th>$r_{\text{cut}}$ (BB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1.245 \times 10^5$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
</tr>
<tr>
<td>2</td>
<td>$5 \times 10^2$</td>
<td>0.2 $\rightarrow$ 2.0</td>
<td>0.2 $\rightarrow$ 2.0</td>
<td>0.2 $\rightarrow$ 2.0</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
</tr>
<tr>
<td>3</td>
<td>$1.25 \times 10^5$</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
</tr>
<tr>
<td>4</td>
<td>$1.25 \times 10^5$</td>
<td>2.0</td>
<td>2.0 $\rightarrow$ 1.0</td>
<td>2.0 $\rightarrow$ 1.0</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
</tr>
<tr>
<td>5</td>
<td>$1.25 \times 10^5$</td>
<td>2.0</td>
<td>0.2</td>
<td>0.2</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
<td>$2.5\sigma$</td>
</tr>
<tr>
<td>6</td>
<td>$5.0 \times 10^5$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>7</td>
<td>$10^4$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
</tbody>
</table>
Table 2.1. Continued

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
<th>$\varepsilon_{AA}$</th>
<th>$\varepsilon_{AB}$</th>
<th>$\varepsilon_{BB}$</th>
<th>$r_{cut}^{(AA)}$</th>
<th>$r_{cut}^{(AB)}$</th>
<th>$r_{cut}^{(BB)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$3.375 \times 10^5$</td>
<td>0.2</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>2</td>
<td>$3.375 \times 10^5$</td>
<td>0.2 → 2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>3</td>
<td>$3.375 \times 10^5$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>4</td>
<td>$10^4$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
</tbody>
</table>

Table 2.2. Lennard Jones parameters for multichain assembly (‘chopped’ 18 mers)

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
<th>$\varepsilon_{AA}$</th>
<th>$\varepsilon_{AB}$</th>
<th>$\varepsilon_{BB}$</th>
<th>$r_{cut}^{(AA)}$</th>
<th>$r_{cut}^{(AB)}$</th>
<th>$r_{cut}^{(BB)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$3.375 \times 10^5$</td>
<td>0.2</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>2</td>
<td>$3.375 \times 10^5$</td>
<td>0.2 → 1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>3</td>
<td>$6.75 \times 10^5$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>4</td>
<td>$3.75 \times 10^6$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>$2.5\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
</tbody>
</table>

2.2 Measures of Shape

In polymer science, the most prevalent measure of shape is the radius of gyration. The squared radius of gyration of a polymer is defined as the mean square average distance of the monomers from the center of mass, $s_i$. For monomers with identical mass:
\[ R_g^2 = \left\{ \frac{\sum_{i=1}^{N} m_i (s_i^2)}{\sum_{i=1}^{N} m_i} \right\} \] \[ = \frac{1}{N} \sum_{i=1}^{N} \langle s_i^2 \rangle \] (12)

Alternative measures of shape involve the radius of gyration tensor. The radius of gyration tensor can be defined as follows:

\[ S = \frac{1}{N} \sum_{i=1}^{N} s_i \cdot s_i^T \] (13)

In three dimensional Cartesian coordinates, diagonalization yields a tensor that can written in terms of the eigenvalues, \( \lambda_i \):

\[ S = \frac{1}{N} \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix} \] (14)

This allows the superimposition of a polymer chain onto an ellipsoid with principal axes proportional to the eigenvalues, wherein \( \lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2 \). Several functional forms of the invariants (I_1, I_2) and eigenvalues of this tensor can serve as shape descriptors for the conformation of the polymer chain(s). This includes:

1. An alternative definition of the squared radius of gyration, \( R_g^2 \), where:

\[ R_g^2 \ = \ I_1 \ = \ \text{tr}(S) \ = \ \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \] (15)

2. The asphericity, b, where:

\[ b = \lambda_3^2 - \frac{1}{2} (\lambda_1^2 + \lambda_2^2) \] (16)

3. The acylindricity, c, where:

\[ c = \lambda_2^2 - \lambda_1^2 \] (17)

4. Aspect ratios:
\[ L_1 = \frac{\lambda_3^2}{\lambda_1^2}, L_2 = \frac{\lambda_2^2}{\lambda_1^2} \] (18)

5. The relative shape anisotropy, \( \kappa^2 \), where:

\[
\kappa^2 = 1 - 3 \cdot \frac{l_2}{l_1^2} = \frac{b^2 + \left(\frac{3}{4}\right) \cdot c^2}{R_g^4} \] (19)

2.3 Clustering

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is utilized to distinguish between nanoglobular conformations through counting the number of solvophobic clusters. A cluster is defined based on a minimum number of points (min_samples) and a specified maximum Euclidean distance (eps) between neighboring points: eps = 1.5 and min_samples = 1.

2.4 Neighbor Count Analysis

For two species A and B, the probability of finding particle B a distance r away from the reference bead A is calculated from the asymmetric radial distribution function:

\[
g_{AB}(r) = \frac{N}{\rho N_A N_B} \sum_{i=1}^{N_A} \sum_{k=1}^{N_B} \langle \delta(r - |r_k - r_i|) \rangle \] (20)

Usually in the determination of the radial distribution function of a bulk system, the system is binned from a reference point (usually the center of mass) across the longest length of the simulation box. Here, to analyze the total number of neighbors for a given bead in a repeating motif, A\sb{\alpha}B\sb{\beta}, the asymmetric radial distribution is analyzed at a cutoff distance, \( r = 1.4 \sigma \) (where \( \sigma = 1 \) in all cases of this work) within a ‘single bin’.
The total number of (solvophobic and solvophilic) neighbors is averaged over the total number of central atoms, \( n \) (the overall sequence is \((A\alpha B\beta)n\)) and the number of trajectory frame/snapshots, \( T \). That is,

\[
\text{Neighbor count} = \frac{\text{Total number of neighbors}}{n \cdot T}
\]  

(21)

\( \text{(a)} \quad \text{(b)} \)

Figure 2.2. Demonstration on the DBSCAN Clustering Algorithm from Scikit-Learn in determining the number of solvophobic clusters in two nanoglobular conformations, (a) string of connected beads, and a (b) toroidal conformation. For each conformation, the top and bottom images represent the system before and after the algorithm detects the clusters. The different colors in the bottom image indicate the number of clusters identified. Here, only solvophobic (A) coordinates are considered for the analysis.
Chapter 3: Design of Single Chain Nanoglobules

3.1 Determination of the Design Space

![Graph showing the number of unique sequences as a function of the length of a binary copolymer.](image)

Figure 3.1. Number of unique sequences as a function of the length of a binary copolymer

The lack of a sequence-shape relationship forces a trial-and-error design approach, which becomes quickly impractical for large molecular weights. For instance, a single binary copolymer chain with a modest molecular weight of 80 already has over $10^{23}$ possible sequences. While there have been significant advances in polymer chemistry to create sequence-controlled polymers\(^1\), synthesis is still a laborious and expensive process.\(^1\)

As highlighted in Figure 3.1, the search space of sequences of a model binary copolymer is massive – a 100 mer copolymer has over $10^{23}$ sequences. This suggests that trial and error search

---

\(^1\) The work in this chapter is based on results submitted to a manuscript in ACS Macromolecules, currently under review.
is intractable; emphasizing the importance of physics-based heuristics, that is, understanding general design rules based on sequence manipulation, as well as, determining a tractable search/design space.

Inspired by this, a genetic algorithm was utilized to provide a guided search of a copolymer sequence that assembles into a targeted nanoglobular shape.

It would be desirable to search for a sequence that collapses into a globular-like structure with high relative shape anisotropy, $\kappa^2$. This sort of shape could be applicable to a cylindrical shape. The design challenge is that coils (copolymer sequences with a large fraction of solvophilic beads) have an inherently large degree of anisotropy; thus a search algorithm would converge on this trivial result.

The nature of this problem thus involves constrained optimization. There are numerous studies that document the use of penalty functions to constrain the design space probed by

![Figure 3.2. (a) Plot of radius of gyration of conformation for homopolymer globules against the reduced temperature of the system. Plot (b) shows the results from a genetic algorithm that ran for 50 generations to search for a copolymer sequence with the maximum anisotropy, highlighting the maximum fitness for each generation. Globular conformations are provided in the inset where beads A and B are colored orange and blue respectively.](image-url)
evolutionary algorithms. To elicit a simple penalty function to screen for ‘globular’ shapes, the coil-to-globule transition of a single homopolymer chain was investigated. This allows one to determine, at what value of $\varepsilon_{\text{HH}}$, the copolymer passed through a coil-to-globule transition.

As highlighted in Figure 3.2, the value of $\varepsilon_{\text{HH}}$ was ramped in 0.2 intervals to 2.0 reduced units. The radius of gyration of the resulting homopolymer globule was averaged after the system was relaxed at a temperature of $T = 1.0$ units. This simulates the decrease in the reduced temperature, $T^* = k_b T/\varepsilon_{\text{HH}}$. According to Figure 3.2, at a reduced temperature corresponding to $\varepsilon_{\text{HH}} = 0.6$ and $R_g \approx 2.98$, the system appears to undergo a coil-to-globule transition plateau. Thus, the function to be maximized can be as follows:

$$F(\text{sequence}) = H[R_g(\text{sequence})] \cdot k^2(\text{sequence})$$

$$H[R_g(\text{sequence})] = \begin{cases} 1, & R_g \leq R_{g,\text{globule}} \\ 0, & R_g > R_{g,\text{globule}} \end{cases}$$

Table 3.1. Genetic algorithm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size</td>
<td>100</td>
</tr>
<tr>
<td>Number of generations</td>
<td>50</td>
</tr>
<tr>
<td>Population size</td>
<td>128</td>
</tr>
<tr>
<td>Selector</td>
<td>Fitness-proportionate</td>
</tr>
<tr>
<td>Crossover</td>
<td>Two-point</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.01</td>
</tr>
</tbody>
</table>

$R_{g,\text{globule}} = 2.98$. Figure 3.2 shows the results obtained for a genetic algorithm that ran for 50 generations. Based on prior work involving a molecular dynamics-simulation based genetic algorithm, the binary copolymer sequence is mapped to a binary genome of 0’s and 1’s.
Each generation evolves via a fitness-proportionate selection operator to select parents; parents undergo recombination via two-point crossover to exchange genetic material, and every genome bit mutates with a given probability. The population size was adjusted each generation to maintain 128 fitness evaluations, wherein repeat candidates are drawn from a census rather than evaluating the fitness again through MD simulations.

Figure 3.3. Conformations of single chain block copolymer globules with sequence, \((AB)_N\) where \(N = 1\) to 50. Beads A and B are colored orange and blue respectively.

From these results, it is not clear what globular conformation the genetic algorithm is converging to. At the very least, the conformation does not exhibit a coil-like conformation – providing evidence that the screening method via the radius of gyration was sufficient to filter the larger solvophilic sequences. Nevertheless, this matrix of over 6400 simulations is computationally expensive and raises the question – what possible nanoglobular shapes are accessible within this bead-spring polymer model? Is full sequence control necessary?
A good starting point to answering these questions is mapping out all possible multiblock copolymers with sequence, \[(AB)^\alpha\] as seen in Figure 3.3, where \(\alpha\) spans from 1 (strictly alternating) to 50 (diblock).

From this qualitative analysis alone, it is evident that most shapes in this design space are sparingly non-spheroidal; bearing resemblance to block copolymer micelles in solution. At the strictly alternating limit, the solvophobic (A) blocks are not sufficiently large enough to force segregation into a particular core area. At the other extreme, the solvophobic and solvophilic blocks are large enough to permit collapse into a tadpole-like structure – with a solvophobic head and a long solvophilic tail.

To access non-trivial shapes not accessible by multimolecular assembly, one must consider probing into the vast design space highlighted in Figure 3.1. Number of unique sequences as a function of the length of a binary copolymer. Motivated by the results for multiblock copolymers...
in Figure 3.3, the design space of all possible 10 mer repeating units was fully mapped and simulated. Omitting the homopolymer coil (all solvophilic beads), this corresponds to 527 total unique sequences. Here, each repeating unit or motif is repeated N times until the desired DOP is reached. Probing accessible shapes at two different degrees of polymerization (DOP) of 100 and 250 yields results as depicted in the plot Figure 3.4. The relationship between the relative shape anisotropy and fraction of solvophobic beads for all 10 mer repeating motifs with overall DOP on left (a) 100 and right (b) 250.

![Figure 3.5](image)

Figure 3.5. Typical globular conformations with repeating motifs (a) A₂B₈, (b) A₄B₆, (c) A₃B₇, (d) A₃B₂, (e) A₄B₂A₂B₂ and (f) A₃B₃A₃B₁ at DOP = 250. Beads A and B are colored orange and blue respectively.

Here, the relationship between the fraction and the relative shape anisotropy of solvophobic (core) beads are plotted. The definition of the relative shape anisotropy is provided in the Methodology. A value of 0 would represent a spherically symmetric system while a value close to 1 would indicate a level of anisotropy wherein the beads tend to align in a straight line. For both
DOPs, sequences that have a large fraction of solvophobic beads appear roughly spheroidal and have a relative shape anisotropy close to 0 – in the limit of a spherical globule.

As shown in Table 3.2 below, 5 different sequences exhibit the necklace conformation. Although these sequences seem different based on the distribution of block (both A and B) lengths, they have a similar ‘end-to-end’ distribution. That is, if one considers the ends of a given sequence, the adjusted block length is a combination of these the ends. For instance, sequence BAAABBBBBB with end blocks: B and BBBBBB can be combined, and thus the ‘equivalent’ sequence is AAABBBBBBB, or simply, A₃B₇. It should be emphasized that this allows one to reduce the design space in a tractable manner, and the sequence of block lengths does not appear to play a significant role in setting nanoparticular shape.

Table 3.2. Repeating sequences of necklace conformations with total length 10 and their respective block length distributions

<table>
<thead>
<tr>
<th>Sequence of repeating motif</th>
<th>Shape of nanoglobule</th>
<th>Block Length Distribution</th>
<th>‘End-to-end’ Block Length Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAAABBBBBB</td>
<td>Necklace</td>
<td>{1, 3, 6}</td>
<td>{3, 7}</td>
</tr>
<tr>
<td>BBBAAABBBBB</td>
<td>Necklace</td>
<td>{3, 3, 4}</td>
<td>{3, 7}</td>
</tr>
<tr>
<td>BBAAABBBBBB</td>
<td>Necklace</td>
<td>{2, 3, 5}</td>
<td>{3, 7}</td>
</tr>
<tr>
<td>AAABBBBBBBB</td>
<td>Necklace</td>
<td>{3, 7}</td>
<td>{3, 7}</td>
</tr>
<tr>
<td>AABBBBBBBBA</td>
<td>Necklace</td>
<td>{2, 7, 1}</td>
<td>{3, 7}</td>
</tr>
</tbody>
</table>
Motivated by these means to make the design space tractable, the genetic algorithm is revisited for the design of an anisotropic globular shape. Here, the sequence of interest is a 25-mer repeat unit with an overall DOP = 250. The same hyperparameters were utilized, however the population size and number of generations were decreased due to limited computational resources. Results are plotted in Figure 3.6 for a total of 1216 sequences. Plot (a) shows the range of anisotropies probed; highlighting that most of the anisotropic sequences are beyond a fraction of solvophobic beads of 0.2. Plots (b) and (c) show the relationship between the standard deviation of block lengths for all sequences, and at a fixed fraction of 0.6 with the relative shape anisotropy.

Figure 3.6. Results from a genetic algorithm to design a repeating sequence that permits folding into an anisotropic globule. Plot (a) shows the relationship between fraction of solvophobic (A) beads and core anisotropy. Plots (b) and (c) show the relationship between standard deviation of block lengths for all sequences, and at a fixed fraction of 0.6 with the relative shape anisotropy.
of the ‘end-to-end’ block lengths and relative shape anisotropy for all sequences, and at a fixed fraction of solvophobic beads of 0.6.

Table 3.3. Top 10 anisotropic sequences from the genetic algorithm

<table>
<thead>
<tr>
<th>Repeating Sequence</th>
<th>Relative shape anisotropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABABBAABBBAAAABBABAAABABA</td>
<td>0.402814</td>
</tr>
<tr>
<td>BAAABABAABBBAAAABBAAAABABAAB</td>
<td>0.387791</td>
</tr>
<tr>
<td>ABAABBABAAAABBBAAAABABABAABABABA</td>
<td>0.387337</td>
</tr>
<tr>
<td>AAABBAABABAAABBBAAABBAABBAABBA</td>
<td>0.386423</td>
</tr>
<tr>
<td>AABBBABABBAABABBAABBAABABBA</td>
<td>0.37551</td>
</tr>
<tr>
<td>ABABBAAABBBAAAABAAABABABAB</td>
<td>0.370272</td>
</tr>
<tr>
<td>ABAABABBAAAABABABABBAAAABBBBA</td>
<td>0.370252</td>
</tr>
<tr>
<td>BBBABBAAAAABBBAAAABBBAAAB</td>
<td>0.366942</td>
</tr>
<tr>
<td>ABABBAAABBBAAAABBBAAABBBABAA</td>
<td>0.365586</td>
</tr>
<tr>
<td>AAABBAABBBAAAABBBAAABBBAAAB</td>
<td>0.362459</td>
</tr>
</tbody>
</table>

From these plots, it is evident that the maximum anisotropic sequences appear to have a very specific block length distribution – with a standard deviation around 1.0. This suggests that a multimodal distribution of blocks if favored. All these sequences have at least a single solvophobic block length, that is, a single A block. These solvophobic (A) singlets might have a propensity to be dragged towards to the interface by larger neighboring solvophilic (B) blocks, in a similar physics seen for in-silico analysis of sequence-specific compatibilizers.35
This is illustrated via analyzing the neighboring counts for each bead in a repeating motif, as illustrated in Figure 3.7. Detailed information on generating the neighboring counts is documented in the Methodology. From this plot, it is evident that neighboring beads influence the position of a bead relative to being in the ‘core’ versus ‘shell’ – being closer to the A/B interface. That is, beads with less neighbors are closer to the interface. For instance, solvophobic beads with indices: 5 and 6, have the lowest neighbor counts out of all solvophobic beads, due to having more solvophilic neighbors.

Overall, results from mapping out all possible 10 mers and utilization of the genetic algorithm for the inverse material design of anisotropic globules highlight some important findings. First, at a sufficiently large molecular weight, the ‘end-effects’ of sequence appear to play no significant role in setting nanoparticular shape. However, the proximity of dissimilar block lengths might play a preferential role in determining the location of a particular bead/segment relative to
the A/B interface. Second, optimizing on relative shape anisotropy alone is a challenge to distinguish between cylindrical and necklace-like conformations.

What anisotropic shapes, other than necklaces and cylinders, are accessible within the design space? Specifically, it is unclear about the shape classification of Figure 3.5 (d). Is it a growing cylinder with a thick radius or a disk-like conformation?

To answer these questions, selected sequences as shown in Figure 3.8 are simulated at higher DOPs of 1000 and 2000. The globular shapes observed are realized through a cascade of increasing shape complexity with increasing molecular weight. Necklace-forming sequences bud into increasing numbers of beads with increasing molecular weight. The shape in Figure 3.5 (d) grows bilaterally into a sheet and ultimately curves into a vesicle when the molecular weight is sufficiently high to allow vesicle formation consistent with the sheet’s bending radius; worm-

<table>
<thead>
<tr>
<th>Repeat Unit</th>
<th>DOP = 100</th>
<th>DOP = 1000</th>
<th>DOP = 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>(ii)</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>(iii)</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 3.8. Effect of degree of polymerization on molecular shape for single copolymer chain globules with repeat units: (I) A₃B₂, (II) A₇B₃ and (III) A₈B₂. A cross-sectional view of each vesicle is shown for DOP = 2000. Beads A and B are colored orange and blue respectively.
forming structures systematically grow in length with increasing molecular weight. This demonstrates that high molecular weights (DOP > 1000) can allow access to complex shapes.

3.2 Effect of Solvophobic Block Length on Nanoglobular Shape

Based on the prior findings, Brownian dynamics simulations were performed using the LAMMPS package on polymer chains with repeating units of the form $A_\alpha B_\beta$ at the large molecular weight limit – DOP $\approx$ 2000. This concept of repeating units/motifs allows a reduction of the design space, allowing one to probe the effect of sequence on molecular shape in a more tractable manner. To gain insight into the possible shapes in this reduced design space, a set of simulations were performed on polymers with repeating motifs ranging from 2-18 segments, that is, from $\alpha + \beta = 2$ to $\alpha + \beta = 18$.

Table 3.4. Repeating motifs of form $A_\alpha B_\beta$ utilized in this work with DOP $\approx$ 2000

<table>
<thead>
<tr>
<th>$\alpha + \beta$</th>
<th>N</th>
<th>DOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>666</td>
<td>1998</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
<td>2000</td>
</tr>
<tr>
<td>5</td>
<td>400</td>
<td>2000</td>
</tr>
<tr>
<td>6</td>
<td>333</td>
<td>1998</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>2000</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>2000</td>
</tr>
<tr>
<td>12</td>
<td>167</td>
<td>2004</td>
</tr>
<tr>
<td>15</td>
<td>133</td>
<td>1998</td>
</tr>
<tr>
<td>18</td>
<td>111</td>
<td>1998</td>
</tr>
</tbody>
</table>
As shown by Figure 3.9, the nanoglobular shapes formed by these high-\(\chi\) sequence-specified chains can be classified into 7 morphological types: expanded coils in which solvophilic interactions dominate the chain morphology; necklaces comprised of solvophobic beads connected by solvophilic tethers; worms (which sometimes curl into tori); sheets (which tend to curl into vesicles with a single solvophilic cavity), essentially homogenous spheroids; sponges containing multiple solvophilic pores; and ill-defined ‘gnarled’ shapes.

The shapes in Figure 3.9 can be rationally accessed based upon the relative values of \(\alpha\) vs \(\beta\). Chains with extremely short solvophobic blocks (low \(\alpha\)) are highly solvophilic and form an expanded coil structure. With increasing \(\alpha\), solvophobic segments assemble into collapsed spheroidal beads connected by coiled solvophilic chains, and this situation persists approximately for \(2 < \alpha < \frac{1}{2}\beta\). When \(\alpha \approx \beta\), the chains form ‘cigar’ or ‘worm’ shapes. For \(\alpha > -2\beta\), sheets are
formed, with these sheets tending to form vesicles. Finally, homopolymers of the solvophobic chain naturally yield filled spheroidal shapes. Gnarled and sponge morphologies are only obtained for very short block lengths.

\[ \theta = \frac{\log(n)}{\log(C)} \]  

Figure 3.10. Quantitative analysis of sequences with repeat unit, \( \alpha + \beta = 10 \): (b) dependence of aspect ratio \( (L_1 = \lambda_3/\lambda_1) \) on solvophobic block length and (c) number of average clusters on solvophobic block length as defined using the DBSCAN clustering algorithm in Scikit-Learn. Lines in all figures correspond to conformation boundaries drawn by visual analysis in VMD. Beads A and B are colored orange and blue respectively.

To distinguish worms, necklaces, and coils from each other (which have a relatively higher degree of anisotropy), an order parameter is defined,

\[ \theta = \frac{\log(n)}{\log(C)} \]  

where \( n \) and \( C \) refer to the number of repeating motifs within a single chain and solvophobic clusters respectively. The number of bead clusters (focusing only on solvophobic beads) are
determined via the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, as explained in the Methodology.

Figure 3.11. Effect of degree of polymerization on molecular shape for single copolymer chain globules with repeat units: (I) A₃B₂, (II) A₇B₃ and (III) A₈B₂. A cross-sectional view of each vesicle is shown for DOP = 2000. Beads A and B are colored orange and blue respectively.

Figure 3.12. Effect of DOP on aspect ratio for three repeating sequences: A₃B₂ (blue circles), A₇B₃ (orange squares) and A₈B₂ (green triangles). Lines are drawn to guide the eye.
Together, a combination of the aspect ratio and cluster analyses can distinguish the various conformations, in agreement with the visual analysis. To illustrate that the minimum molecular weight to achieve non-spherical shapes is sequence-dependent, qualitative, and quantitative descriptions are provided for three repeating motifs/sequences that form sheets/vesicles: A₃B₂, A₇B₃ and A₈B₂.

To quantify asphericity, the aspect ratio, $L_1$ is utilized. As a globule grows into a sheet with increasing degree of polymerization (DOP), the aspect ratio increases. Ultimately, the sheet folds into a vesicle and the aspect ratio dramatically decreases; wherein the thickness of a sheet determines the solvophile concentration in the core. The cascade of these events is sequence-dependent as shown quantitatively in Figure 3.12 and visually in Figure 3.11. In Figure 3.12, eigenvalues of the radius of gyration tensor (involving only solvophobic beads) are calculated in LAMMPS and averaged over a 2 million timestep production period.

As can be seen here for three representative sequences coding for sheets, sequences with lower $\alpha$ exhibit a dramatic increase in aspect ratio at lower degree of polymerization than do chains with higher $\alpha$. This indicates a transition from spheroid to sheet at lower molecular weight.

Figure 3.13. Number density ($\rho$) binned from the center of the mass ($r$) for macromer, A₃B₂. Wall thickness, $d_{wall}$ and cavity radius, $r_{cav}$ are determined as highlighted in the plot.
In all cases the aspect ratio ultimately drops at high molecular weight – this is a consequence of the sheet curving into a vesicle at high degree of polymerization.

To characterize the dimensions of single chain vesicles – wall thickness and cavity radius – the number density profile, binned from the radial distance relative to the center of mass, is analyzed as highlighted in Figure 3.13. The wall thickness is calculated as the difference in radii at the half the maximum density of solvophobic beads (species A), while the cavity thickness is defined as the radial distance at the inner of these two points.

![Figure 3.13](image)

**Figure 3.14.** Dependence of vesicle (i) wall thickness and (ii) solvophilic cavity radius on solvophobic block length ($\alpha$) for solvophilic block lengths ($\beta$) ranging from 1 to 4 as indicated in the inset.

As shown in Figure 3.14, larger solvophilic blocks produce thicker vesicle walls, which for a given molecular weight naturally yield smaller pores; conversely, larger solvophilic blocks pack the solvophilic core more densely and thus expand the cavity.
This level of cavity size control is qualitatively distinct from the situation in multi-molecular vesicles such as BCP and lipid vesicles, where control is limited by the fact that the cavity size is intrinsically controlled by the number of molecules per vesicle, which is not a directly controllable quantity. In this case, the equivalent to molecules per vesicle is the chain molecular weight, which is amenable to direct control, such that at fixed molecular weight sequence effects can dominate.

![Diagram](image)

**Figure 3.15.** Non-normalized radial distribution function plots for sequences with repeating motifs, $A_\alpha B_\beta$, for (i) $\alpha + \beta = 6$, (ii) $\alpha + \beta = 8$, (iii) $\alpha + \beta = 10$, and (iv) (i) $\alpha + \beta = 12$.

As seen in Figure 3.15, the non-normalized radial distribution is plotted for various sequences with repeating motif, $A_\alpha B_\beta$, for (i) $\alpha + \beta = 6$, (ii) $\alpha + \beta = 8$, (iii) $\alpha + \beta = 10$, and (iv)
\[ \alpha + \beta = 12. \] The non-normalized radial distribution or coordination number is defined as:

\[ n(r) = 4\pi \rho \int_0^r g(\zeta)\zeta^2 d\zeta \]  

(25)

where \( g(r) \) is the radial distribution function, defining the probability of finding a particle at a distance \( r \) from a reference particle. \( \rho \) is the number density, \( \frac{N}{V} \).

One can ascertain various length scales and thus the dimensions of the nanoparticular shapes from these plots. Coils (sequences with \( \alpha \leq 2 \)) have the lowest coordination number or \( n(r) \) based on the lack of any short-range order as seen in other systems such as cylinders and necklace conformations. This opens the opportunity for experimental validation as the radial distribution function can be ascertained from multiple characterization methods.

### 3.3 Effect of Interfacial Length on Nanoglobular Shape

The resulting conformations and conformation boundaries depicted in Figure 3.3 are strongly dependent on the interfacial separation between solvophobic and solvent beads – which in this coarse-grained model, relates to the self-attractive Lennard Jones interaction parameter, \( \varepsilon_{AA} \).

As shown by Figure 3.16, the conformation diagrams at \( \varepsilon = 1.5 \) is qualitatively like \( \varepsilon = 2 \), with a greater proportion of coil and necklace conformations. At \( \varepsilon = 1 \), the AB segregation is not strong enough to force collapse into definite shapes for many sequences - worm, sheet and vesicle conformations almost disappear. These results indicate that strong solvent-quality contrast can yield complex shape formation.

### 3.4 Effect of Stiffness on Nanoglobular Shape

In the simplest model, a polymer conformation with \( n \) bonds and bond length \( l \) is based on a random walk with a mean end-to-end squared distance, \( nl^2 \). Real polymer conformations,
however, do not behave this way, due to bond, angle, dihedral and other restrictions. Nevertheless, the basis of coarse-grained models allows one to map an arbitrary polymer conformation to a random-walk chain with mean-end-to-end squared distance,

\[ \langle h^2 \rangle_0 = C_\infty n l^2 = N b^2 \]

where, \( C_\infty \) is the characteristic ratio and quantifies the effect of local restrictions/constraints; \( N b^2 \) describes the ‘mapped’ random-walk conformation with \( N \) and \( b \) representing the number and length of connected Kuhn segments respectively.

In coarse-grained bead-spring polymer models, the Kuhn length can be varied through the addition of a cosine bending potential:

\[ E_{\text{angle}} = K_\theta (1 + \cos \theta) \]

where \( K_\theta \) is the stiffness or bending constant. As seen in Figure 3.17 and Figure 3.18, \( K_\theta \) is varied from 1.5, 3.0 to 4.5 and conformations are visualized for 5-mer repeating motifs.

In Figure 3.17, all conformations correspond to \( \epsilon = 2.0 \), like that employed in Figure 3.3. For a given sequence, moving left to right, one can ascertain the effect of stiffness on the resulting nanoparticular shape. Repeating motif, \( A_4B_1 \), appears a ‘sponge’ morphology, wherein there a non-localized solvophilic cavity at \( K_\theta = 0 \) (seen in Figure 3.3). The effect of stiffness appears to localize these areas into stripes. Also, it should be noted that the overall nanoparticular shape moves from being roughly spherical to a more rigid or box-like structure. From visual analysis alone, it appears that the system is crystallizing at \( K_\theta = 4.5 \). Closer inspection of the solvophobic beads confirms that the beads are forming facets. This is also seen for the other sequences. An interesting result is seen for sequence, \( A_3B_2 \), which is unable to form sheets/vesicles in this regime. At \( \epsilon = 1.5 \), the effect of stiffness appears to pin the edges, and the conformation resembles a sandwich-like or
hairpin structure. Increasing stiffness appears to produce multiple stacked layers, like a lamellar configuration.

At a reduced interfacial length set by $\epsilon = 1.5$, similar results are seen as dictated by Figure 3.18; however, from visual analysis, it appears that no system has crystallized. Further quantitative analysis involving the use of a bond autocorrelation function or structure factor might be able to ascertain whether the system is crystallizing. For instance,

$$\langle \hat{r}_i \cdot \hat{r}_0 \rangle = \exp\left(-\frac{i\sigma}{l_p}\right)$$

(28)

Here, $\hat{r}_i$ and $\hat{r}_0$ refer to unit vectors along bond $i$, and bond 0 (the first bond) respectively. $l_p$ is the persistence length, and $\sigma$ refers to the interatomic distance between beads.

Figure 3.16. Conformation diagram of single chain copolymer globules with repeat unit, $A_\alpha B_\beta$ at DOP = 2000 at varying A-B attractive interaction: (i) $\epsilon = 1$, (ii) $\epsilon = 1.5$ and (iii) $\epsilon = 2$. 

47
Figure 3.17. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 5$ at $\epsilon = 2.0$. A cross-sectional view of some conformations is shown on the right. Beads A and B are colored orange and blue respectively.

Figure 3.18. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 5$ at $\epsilon = 1.5$. A cross-sectional view of some conformations is shown on the right. Beads A and B are colored orange and blue respectively.
Overall, it appears that the effect of bond stiffness drastically changes the classes of nanoparticulate shapes accessed in comparison to the results seen in Figure 3.3. Additional conformations are visualized for 10-mer repeating motifs in the Appendices.

3.5 Comparison to Multimolecular Assembly

Given the evidence described above for potentially unique advantages of single molecule assembly, one might ask to what extent these results differ from multimolecular assembly of the underlying motif unit $A_\alpha B_\beta$ comprising the chain. As shown in the Appendices, a test of this question for length-18 motifs (our largest motifs) indicates that, when not concatenated into a single molecule, these motifs generally do not assemble into single globules in simulation even over an annealing period of roughly 7.5 times that provided to the single molecules. Assembly is further attempted by temporarily subjecting these multimolecular systems to a weak drag force towards one another. As shown in Figure 3.19, this induces assembly of distinct chain into single globules when $\alpha \geq 14$. However, unlike the corresponding single-chain globules, which yields vesicles, the resulting globules are micelles lacking an internal cavity. Moreover, motifs that code for worms and necklaces when incorporating into single molecules, as well as $\alpha < 14$ motifs that code for vesicles, do not assemble into single globules even with this drag force.

It is likely that different results would be achieved for multi-molecular assembly if the chains were disassembled into larger and more complex multiblocks rather than their base motif. However, at a minimum, these results indicate that long-range chain connectivity has a quantitative, and even qualitative, effect on the resulting assembled structure. Indeed, existing theories omit several features that are important to the physics of shape in single-molecule assembly and likely account for the differences from multimolecular assembly.
It should be emphasized that the level of cavity size control in the single chain vesicles depicted in Figure 3.3 is qualitatively distinct from the situation in multi-molecular vesicles, where control is limited by the fact that the cavity size is intrinsically controlled by the number of molecules per vesicle, which is not a directly controllable quantity. In this case, the equivalent to molecules per vesicle is the chain molecular weight, which is amenable to direct control, such that at fixed molecular weight sequence effects can dominate.

Notwithstanding, there are similar results of nanoparticular shapes from in-silico and experimental works involving multimolecular assemblies.\textsuperscript{21,92} For instance, previous work\textsuperscript{92} involving DPD simulations of multimolecular assembly of ‘alternating-structured copolymers’ consisting of repeat unit, \((A_xB_y)_n\), (where A and B are solvophilic and solvophobic respectively, in an opposite convention of this work) has shown that the feature sizes of the aggregates from alternating structured copolymers are thinner and more uniform compared to block copolymer assemblies. Transitions in polymer conformation of these alternating structured copolymers parallel some of the results presented in this manuscript – spheroids grow into bead-necklace conformations (micelle networks), worm-like micelles or vesicles with increasing degree of polymerization.

The conformation diagram seems consistent with results predicted for vesicles. For instance, multimolecular assembly of copolymers with repeating motifs, \(A_5B_{10}\) and \(A_4B_{12}\) produce vesicles. This is in general agreement with the conformation diagram in Figure 3.3, with vesicles in the range \(\alpha > -2\beta > 0\).

### 3.6 Elucidating a Sequence-Shape Relationship

How can a rational description of the effect of sequence on nanoglobular shape be presented? A good starting point is understanding the important factors identified in the shape
transformations of block copolymer micelles in solution. As mentioned in the Background and Literature Review, Nagarajan moves from earlier theoretical descriptions of block copolymer micelles. In Nagarajan’s model, the underlying shape of a block copolymer micelle is dependent on transfer free energies associated with the transfer of solvophobic segments towards a ‘core’ area and away from the solvent, the free energy associated with creating an unfavorable (A/B) interface, osmotic swelling of solvophilic segments, and the elastic deformation of all (solvophobic and solvophilic) segments.

Within a single chain made of multiple blocks of solvophobic and solvophilic segments connected, these forces are expected to play a role in setting nanoglobular shape. However, the single-chain connectivity introduces an additional physics into the system, and the challenge of comparing small (multimolecular) molecule assembly versus single chain folding in solution. Indeed, in the prior section, it was shown that wherein a long single chain is ‘chopped’ into its smaller repeating segments, the resulting assembled shape differs in dimensional control (for instance, the solvophobic cavity size). These results indicate that single-molecule assembly can yield qualitatively different structures than assembly of the corresponding continent small molecules – a result of the additional chain connectivity induced in the single molecule.

Entropic contributions due to long chain behavior and the presence of loops (as seen in other works studying interfacial behavior of sequence-controlled polymers may contribute to this differences and ultimately advance the theoretical development in elucidating a sequence-shape relationship.

Nevertheless, one can rationalize the transitions in the conformation diagram as follows. One can imagine mentally walking up a particular straight line along Figure 3.9, for instance, \( \alpha + \)
\( \beta = 10 \). For \( \alpha = 10 \ (\beta = 0) \), the system behaves a homopolymer in a poor solvent. The interfacial cost with creating an unfavorable (A-S where S is the ‘solvent’) interface forces the collapse into a shape with the minimal surface area – a sphere. Moving along the line, the addition of solvophilic segments introduce a transfer free energy contribution. The movement of solvophilic segments towards the surface and away from the solvophobic segments increases the overall curvature and interfacial area. The system responds by curving bilaterally into a sheet. As the solvophobic fraction further increases, osmotic dilution and the elastic deformation of these segments’ forces collapse into a shape with a larger surface area - a worm. Eventually, the solvophilic segments are so large (beginning at \( \alpha = 4 \) and \( \beta = 6 \)) that the growing repulsion between the segments forces the worm to ‘break’ and transition a system of interconnected beads or a necklace conformation. As the solvophilic block gets longer, these beads grow in number, decrease in size. Ultimately, the limit of a polymer in a good solvent (\( \alpha = 0, \beta = 10 \)) is reached, resulting in a coil-like conformation.

### 3.7 Summary and Conclusions

In this first part of the work, Brownian dynamics simulations were used to develop design rules for single chain nanoparticular shapes based on purely non-specific solvophobic/solvophilic interactions. The design space of sequence-controlled polymers is massive, and search (trial and error and guided by genetic algorithms) did not yield any non-spheroidal shapes at degrees of polymerization lower than 250.

At such higher degrees of polymerization, sequence control was investigated at a repeat unit level (from 2 to 25). Analysis involving the non-normalized radial distribution (number density) highlighted that end effects are not as important; the solvophobic/solvophilic block length is vital in setting nanoparticular shape. In a demonstrated case of single chain vesicles, larger
solvophobic block lengths yielded vesicles with thicker walls and smaller degrees of solvophobic cavitation.

Investigation of the effect of interfacial length (indirectly based on changing the solvent quality) demonstrated that strong solvent quality contrast is necessary for complex shape formation. At lower values, sheet/vesicle and worm vesicles progressively disappear, and shape classification becomes a difficult task. Increasing the stiffness of bonds tends to increase the structural ordering, as seen in sponge and vesicle morphologies. Moreover, a completely new family of shapes are accessible; vesicles with solvophobic cavitation are not seen as in flexible system. All systems appear to crystallize at $\epsilon = 2.0$.

The path towards a sequence-shape relationship is still a challenge. Analogies were presented based on extensions of the de Gennes model proposed by Nagarajan. Factors specific to single chains such as long-range connectivity still need to be accounted for.

Finally, the single chain nanoparticular shapes were compared to results achieved from multimolecular assembly. Similar shapes such as worm-like micelles and vesicles can be achieved via the multimolecular assembly of amphiphilic multiblock copolymers in solvent. In the case of single chain nanoparticular shapes, dimensional control is linked to sequence versus aggregation number (an equilibrium quantity) in multimolecular assembly.
Figure 3.19. Comparison of conformations of multimolecular assemblies (with a drag force of 0.01 LJ units) with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$. Assembled structures have a cross-section at the right. Beads A and B are colored orange and blue respectively.
Chapter 4: Design of Aggregation-Resistant Single Chain Vesicles

4.1 Introduction

Equipped with a set of design rules for designing single chain nanoparticular shapes, one might ask how do these objects interact in solution? That is, do these globules act independently or aggregate with increasing concentration?

Synthesis at low concentration is generally a viable approach; however, actual usage of these shape-specified polymer nanoparticles will require them to be reasonably shape-stable at higher concentrations, and thus act as ‘single chain technologies’.95 Post-processing techniques exist that can lock into the desired shape such as crosslinking or vitrification. One potential strategy for shape stabilization, inspired by stabilization of nanoparticles against aggregation via polymer grafting96, is to incorporate into the chain long solvophilic segments intended to form a halo around the polymer nanoparticle.

Here, the focus is on single chain vesicles depicted in Figure 3.3.2 Various iterations involving the attachment of solvophilic loops or halos within the sequence are incorporated to enable aggregation resistant at low concentration. The design challenge involves designing a system of sequence-specified vesicles that resistant aggregation at low concentration with minimal alteration to the underlying shape of the single-chain vesicle.

---

2 The work in this chapter is based on results submitted to a manuscript in ACS Macromolecules, currently under review
4.2 Methodology

First, the multichain assembly of nanoglobular shapes with repeating motifs and total DOP of 2000 is simulated. For each system, five copolymer chains start in the globular configuration (coordinates were retrieved from the annealed simulations as depicted in Figure 3.3), and a drag force is similarly employed. It is then turned off and the system is analyzed after an additional annealing period. Second, the aggregation behavior at finite concentration of single chain vesicles (‘grafted’ and bare) is investigated. These simulations contain 20 single chain vesicles. Each vesicle is initialized based on coordinates obtained from annealed single-chain simulations as described above. To accelerate the assembly process, a *smaller drag force (0.001 LJ units)* is employed that drags all molecules towards their common center of mass. As shown in the table below, the assembly process is simulated for 250,000 τ and track the number of solvophobic clusters over time based on the clustering algorithm defined earlier. Results are averaged over three different trials.

Table 4.1. Lennard Jones parameters for multichain assembly of 2000 mer globules

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration (τ)</th>
<th>ε_AA</th>
<th>ε_AB</th>
<th>ε_BB</th>
<th>r_cut (AA)</th>
<th>r_cut (AB)</th>
<th>r_cut (BB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0125 × 10^6</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5σ</td>
<td>2^{1/6}σ</td>
<td>2^{1/6}σ</td>
</tr>
<tr>
<td>2</td>
<td>3.375 × 10^5</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5σ</td>
<td>2^{1/6}σ</td>
<td>2^{1/6}σ</td>
</tr>
<tr>
<td>3</td>
<td>3.75 × 10^6</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5σ</td>
<td>2^{1/6}σ</td>
<td>2^{1/6}σ</td>
</tr>
</tbody>
</table>
4.3 Aggregation Behavior of ‘Bare’ Nanoglobular Shapes

How stable are the shapes achieved in single-molecular assembly against aggregation at finite concentration?

Figure 4.1. Aggregation behavior of single chain copolymer globules with repeating motifs: (a) A3B2, (b) A3B7 and (c) A5B5 for simulation protocols without and with a drag force. Each assembly consists of 5 copolymer chains initially in the globular configuration with DOP = 2000. A force of 0.01 LJ units, as seen on the far right for each globule, is utilized to drag all atoms to their common center of mass. A cross-sectional view of the resulting multi-layered vesicle is shown for (a) on the right. Simulation images were rendered in VMD.

To answer this, assemblies of multiple single-chain copolymer globules encoding sheet/vesicle, necklace and worm conformations were simulated. As shown by the snapshots in Figure 4.1 reflecting the system state after approximately ten times the annealing time employed to produce the original globules, these globules can in some cases aggregate, although they do so quite slowly. After this time, the worms remain as separate globules; necklaces and vesicles exhibit modest leavening. As with multi-molecular assembly, behavior in the presence of a drag force...
between molecules favoring assembly is also investigated. In this case, vesicles assemble into an ‘onion’ structure as seen in other work\(^5\). It is not clear whether this is an equilibrium state or is simply kinetically trapped after being favored during application of the drag force. Necklaces appear to fully intertwine. Worms remain relatively resistant to assembly – a likely consequence of torus-formation stabilizing the single-chain state. This also may parallel results in semi-flexible polymers and DNA condensates\(^{39,97,98}\), where formation of a single large toroid is gated by a large kinetic barrier.

4.4 Design of Aggregation-Resistant Vesicles

To probe the effect of adding solvophilic segments into the single chain vesicle, various lengths and number of repeating motifs were tested as seen in the table below. Resulting conformations of 6 out of the total 7 vesicles are visualized in Figure 4.2.

Table 4.2. Sequences of grafted single chain vesicles with vesicle motif, A\(_3\)B\(_2\)

<table>
<thead>
<tr>
<th>Vesicle Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B_{200}(A_3B_2)_{400})</td>
</tr>
<tr>
<td>(B_{100}(A_3B_2)<em>{400}B</em>{100})</td>
</tr>
<tr>
<td>([B_{50}(A_3B_2)_{133}]_3)</td>
</tr>
<tr>
<td>([B_{25}(A_3B_2)_{100}]_4)</td>
</tr>
<tr>
<td>([(A_3B_2)<em>{40}B</em>{50}]<em>9-(A_3B_2)</em>{40})</td>
</tr>
<tr>
<td>([(A_3B_2)<em>{40}B</em>{75}]<em>9-(A_3B_2)</em>{40})</td>
</tr>
<tr>
<td>([(A_3B_2)<em>{40}B</em>{100}]<em>9-(A_3B_2)</em>{40})</td>
</tr>
</tbody>
</table>
Specifically initial focus is on the vesicle motif, $A_3B_2$, which is the thinnest sheet/vesicle and with the largest degree of solvophobic cavitation. ‘Grafted’ vesicle (i) exhibits a long solvophilic tail tethering off the parent vesicle. From the cross-sectional view, it appears that the vesicle cavitation is unaltered.

![Vesicle Motif](image)

Figure 4.2. Vesicles with sequence, (i) $B_{200}(A_3B_2)_{400}$, (ii) $B_{100}(A_3B_2)_{400}B_{100}$, (iii) $[B_{25}(A_3B_2)_{100}]_4$, [(A$_3$B$_2$)$_{40}$B$_y$]$_9$-(A$_3$B$_2$)$_{40}$ with (iv) $y = 50$, (v) $y = 75$, and (vi) $y = 100$. Images were rendered in VMD. Beads A (solvophobic) and B (solvophilic) are colored orange and blue respectively. Cross-sectional views of each vesicle is shown in the middle. A rendered image without the solvophilic beads is shown for each vesicle on the far right.

Addition of repeating solvophilic segments produces loops as seen for vesicles (ii) and (iii). Finally, vesicles (iv) through (vi) have the sequence [(A$_3$B$_2$)$_x$B$_y$]$_{n-1}$-(A$_3$B$_2$)$_x$. Here the A$_3$B$_2$ unit is the repeating vesicle-coding motif and the B$_y$ unit is the solvophilic loop. Three cases are considered: $x = 40$ and $n = 10$, with three solvophilic loop lengths: $y = 50$, $y = 75$ and $y = 100$. It should be noted that there is a missing B$_y$ unit at the end of the sequence. This allows the formation of a fixed number of solvophilic loops around the vesicle, without any dangling ends as seen for vesicle (i). Although vesicles (iii) through (vi) produce a halo of solvophilic loops around the...
vesicle, the shape of the underlying vesicle is altered as depicted from the solvophobic (only) rendering of the conformations in Figure 4.2. Specifically, the solvophilic chains protrude from the vesicle wall on the inner leaflet of the vesicle; essentially poreating the vesicle creating a set of holes. The solvophilic chains pay a significant energetic cost to avoid an even larger entropic cost associated with confinement in the vesicle. If one wishes to stabilize the vesicle without perturbing its shape, this must be avoided. To avoid this poreation, each solvophilic segment has an attached solvophobic segment that biases these loops towards the outer extremities of the vesicle, that is, the vesicle’s surface. A larger solvophobic block compared to the bare vesicle repeating motif should favor movement in the outer leaflet. The resulting sequence is \([A_6B_2]^x A_2B_y]_{n-1}(A_6B_2)^x A_z\) and seven different combinations are presented in Table 4.3. In the trivial case of the bare vesicle, \(n = 24, x = 1, y = 0, z = 0\).

Table 4.3. Vesicles studied at low concentration with sequence, \([A_6B_2]^x A_2B_y]_{n-1}(A_6B_2)^x A_z\)

<table>
<thead>
<tr>
<th>Label</th>
<th>(n)</th>
<th>(x)</th>
<th>(y)</th>
<th>(z)</th>
<th>(\langle n_{A, end} \rangle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>24</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>(ii)</td>
<td>9</td>
<td>25</td>
<td>50</td>
<td>9</td>
<td>3.33</td>
</tr>
<tr>
<td>(iii)</td>
<td>9</td>
<td>25</td>
<td>50</td>
<td>9</td>
<td>1.88</td>
</tr>
<tr>
<td>(iv)</td>
<td>24</td>
<td>10</td>
<td>20</td>
<td>9</td>
<td>1.33</td>
</tr>
<tr>
<td>(v)</td>
<td>24</td>
<td>10</td>
<td>40</td>
<td>9</td>
<td>1.03</td>
</tr>
<tr>
<td>(vi)</td>
<td>24</td>
<td>10</td>
<td>60</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>
Here, focus is on the vesicle motif, $A_6B_2$, (40% thicker than the vesicle, $A_3B_2$, as seen in Figure 3.14). Here the $A_6B_2$ unit is the repeating vesicle-coding motif and the $B_y$ unit is the solvophilic loop, as above. The new $A_9$ unit is chosen to be 50% longer than solvophobic block in the primary vesicle-coding motif, such that it should favor the outer leaflet. As shown in Figure 4.3, this approach indeed eliminates the poreation problem by placing the long loops reliably on the outer leaflet.

4.5 Aggregation Behavior

Using the ability to place stabilizing loops on the outer leaflet via the strategy above, the aggregation behavior of the vesicles in Table 4.3 at low concentration.

Figure 4.3. Simulation images, rendered in VMD, are shown in (b) for vesicle with repeating motif (i) $A_6B_2$, and $[(A_6B_2)_x A_9(B_y)]_{n-1}(A_6B_2)_x A_9$ for (ii) $x = 25$, $y = 50$, (iii) $x = 25$, $y = 75$, (iv) $x = 10$, $y = 20$, (v) $x = 10$, $y = 40$, and (vi) $x = 10$, $y = 60$. Beads A (solvophobic) and B (solvophilic) are colored orange and blue respectively. Cross-sectional views of each vesicle are shown in the middle.
As seen in Figure 4.4, the system of vesicles without any grafted solvophobic chains, $A_6B_2$, aggregates into one cluster after approximately $50,000\ \tau$. A mean aggregation number is defined as,

$$
\langle n_{A,\text{end}} \rangle = \frac{N_{A0}}{N_{A,\text{end}}} 
$$

(29)

where $N_{A0}$ and $n_{A,\text{end}}$ refer to the number of total unmerged vesicles/clusters at the start and at $250,000\ \tau$ respectively. As outlined in the methodology, $n_{A0} = 20$. The solvophilic blocks form loops on the vesicle’s surface, promoting steric repulsion between chains on neighboring vesicles.
The greatest resistance to aggregation is seen for the vesicle with the greatest density (24) of solvophilic chains and the largest DOP of 60.

As in polymer grafted nanocomposites these ‘self-grafted’ vesicles evidently stabilize the particles against aggregation. In so doing, they stabilize particle shape by preventing the solvophobic cores of the vesicles from ever coming into contact. As a reference\(^9^9\), the aggregation of nanoparticles can be modeled as follows:

\[
\frac{dC_i}{dt} = \frac{1}{2} \sum_{j=1}^{i-1} k_{j,i-1} C_j C_{i-1} - \sum_{z=1}^{\infty} k_{iz} C_i C_z \quad \text{for } 1 \leq i < \infty \quad (30)
\]

\[
\frac{dC_z}{dt} = -\sum_{z=1}^{\infty} k_{1z} C_i C_z \quad \text{for } i = 1 \quad (31)
\]

where \(C_i\) is the concentration of an aggregate containing \(i\) monomers, \(k_{iz}\) is the dissociation rate constant, and other rate constants refer to aggregation for a given number of monomers. At short times, the aggregation can be modeled as a dimer association/dissociation process with aggregation and dissociation constants, \(k_{11}\) and \(k_{12}\) respectively. For irreversible aggregation, the equations decompose to:

\[
\frac{dC_2}{dt} = \frac{1}{2} k_{11} C_0^2 \quad (32)
\]

The above equation is however not exactly suited for aggregation for (nano) particles that do not have a polymeric structure. Moreover, the soft nature or the mechanical response of the chains within solution would need to be accounted for.
If the aggregation behavior was modeled by a single activated process, one can fit the data in Figure 4.4 to an exponential function as follows:

\[ N(t) = (N_0 - N_{\infty}) \cdot e^{-\left(\frac{t}{\tau}\right)} + N_{\infty} \]  

(33)

Where \( N_0 \) and \( N_{\infty} \) refer to the number of aggregates at time 0 (in this work, \( N_0 = 20 \)), and at equilibrium respectively. Here, \( N_{\infty} \) and \( \tau \) are fitting parameters. As seen in Figure 4.5, the exponential fit does an average fit over the entire timescale probed, however, generally overestimating aggregation at small timescales. This suggests that the aggregation process is a
multi-step process, likely with an incubation period at small timescales. The fitting parameters are presented in the Appendices.

Figure 4.6. Aggregation behavior of vesicle with repeating motif, A₆B₂, in the absence of a drag force, via tracking the number of solvophobic clusters over time as defined using the DBSCAN clustering algorithm in Scikit-Learn.

Overall, the increased aggregation timescale of the weakest stabilized vesicles shown in Figure 4.4 indicates that the loops at a minimum provide significant kinetic stabilization. The emergence of a plateau is suggestive of thermodynamic stabilization against aggregation. However, this cannot be determined with certainty, particularly given the limited timescales accessible to simulation. Particle stability against aggregation can be expected to be still higher in the absence of a drag force. The aggregation behavior of the system of vesicles with repeating motif, A₆B₂, without a drag force as highlighted in Figure 4.6. During the same simulation time, most of the vesicle remain unmerged.

Overall, the ‘polymer halo’ effect employed to stabilize grafted nanoparticles can be directly employed in sequence-specified chains. This provides a pathway towards stabilizing particle shape in relatively concentrated solution after initial synthesis and assembly in dilute
solution. Further stabilization could then be provided via vitrification or crosslinking as in some prior works.

4.6 Summary and Conclusions

In this part of the work, single chain vesicles were designed for aggregation resistance at low concentration. Multimolecular assembly of typical vesicles, worms and necklaces appear to aggregate at low concentration, except for worms which tend to independently curl into toroids rather assembling into a large worm/toroid.

To engineer single chain nanoparticles resistant against aggregation, long solvophilic chains were incorporated into the underlying sequence. This was motivated by grafting in inorganic nanoparticle system and polymer nanocomposites, where steric repulsion of polymer chains at the exposed surface prevents aggregation. Direct attachment of the solvophobic chains into sequence result in poreation of the vesicle across multiple sections. To avoid this poreation, longer solvophobic placer segments were placed into the sequence. The engineered sequence was designed, \([(A_{\alpha}B_{\beta})_x A_z B_y]_{n-1}-(A_{6}B_{2})_x A_z\), where \(A_{\alpha}B_{\beta}\) is the underlying vesicle motif and the placer segment, \(A_z\), \(z > \alpha\).

The aggregation behavior of these ‘grafted’ vesicles was then investigated in the presence of a small drag force to accelerate the assembly process. These simulations are computationally expensive; the total timescale probed for each vesicle is around 250,000 \(\tau\). 100 % resistance to aggregation is seen for the vesicle with the greatest density (24) of solvophilic chains and the longest chain, with a DOP of 60. Finally, the kinetics of aggregation in these grafted vesicles still needs to be ascertained. Further analysis could involve modeling the aggregation as a multi-step process with an initial incubation period wherein neighboring globules need sufficient time to diffuse towards each other to aggregate.
Chapter 5: Design of Motif Multiblock Copolymer Globules

5.1 Introduction

Up to this point, the structures obtained in Figure 3.3 mostly parallel those accessible via block copolymer assembly, albeit with a higher level of dimensional control and stabilization against aggregation via sequence-specific design of single chain vesicles. The one exception is the case of the bead-necklace structure; the analogous BCP structure would simply be distinct micelles. Could the use of large single-molecular assembly in the presence of only nonspecific solvophobic/solvophilic interactions allow access to more complex structures as in proteins?3

The single chain nanoparticular shapes designed in this work bear close analogs to the primary structure of proteins – wherein amino acids (or peptides) are polymerized in a sequence-specific manner to produce a polypeptide. To design a hierarchy of structures, noncovalent interactions between amino acids in the polypeptide(s) are involved.100,101 Specifically, interactions such as hydrogen bonding lead to the secondary structure, and can produce various structures such as α helices and β sheets. Tertiary structures are possible through a multitude of additional interactions such as (the poorly understood) hydrophobic interactions, leading to a three-dimensional structure. For systems involving more than one polypeptide chain, additional noncovalent interactions between different polypeptide chains led to more complex structural control, for instance, as seen in the structure of hemoglobin102.

---

3 The work in this chapter is based on results submitted to a manuscript in ACS Macromolecules, currently under review
5.2 Methodology

5.2.1 Brownian Dynamics

The Brownian dynamics simulations presented in the section have the similar protocol for both thermodynamic routes as seen in Table 2.1, with longer simulation times to account for the larger DOPs in the motif diblock copolymers. Table 5.1 below highlights the Lennard Jones parameters and the corresponding simulation timescales for each step in these protocols.

Table 5.1. Lennard Jones parameters for thermodynamic routes for motif diblocks

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration (τ)</th>
<th>ε_AA</th>
<th>ε_AB</th>
<th>ε_BB</th>
<th>r_cut (AA)</th>
<th>r_cut (AB)</th>
<th>r_cut (BB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5 × 10^5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>2.5σ</td>
<td>2.5σ</td>
<td>2.5σ</td>
</tr>
<tr>
<td>2</td>
<td>5 × 10^2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>2.5σ</td>
<td>2.5σ</td>
<td>2.5σ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td>→ 2.0</td>
<td>→ 2.0</td>
</tr>
<tr>
<td>3</td>
<td>2.50 × 10^5</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.5σ</td>
<td>2.5σ</td>
<td>2.5σ</td>
</tr>
<tr>
<td>4</td>
<td>2.50 × 10^5</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.5σ</td>
<td>2.5σ</td>
<td>2.5σ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>→ 1.0</td>
<td>→ 1.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5 × 10^5</td>
<td>2.0</td>
<td>0.2</td>
<td>0.2</td>
<td>2.5σ</td>
<td>2.5σ</td>
<td>2.5σ</td>
</tr>
<tr>
<td>6</td>
<td>7.5 × 10^6</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5σ</td>
<td>2^{1/6}σ</td>
<td>2^{1/6}σ</td>
</tr>
<tr>
<td>7</td>
<td>5.0 × 10^5</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5σ</td>
<td>2^{1/6}σ</td>
<td>2^{1/6}σ</td>
</tr>
</tbody>
</table>
Table 5.1. (continued)

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
<th>$\epsilon_{AA}$</th>
<th>$\epsilon_{AB}$</th>
<th>$\epsilon_{BB}$</th>
<th>$r_{cut}$ (AA)</th>
<th>$r_{cut}$ (AB)</th>
<th>$r_{cut}$ (BB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$6.25 \times 10^5$</td>
<td>0.2</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5$\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>2</td>
<td>$6.25 \times 10^5$</td>
<td>0.2 $\rightarrow$</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5$\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$7.5 \times 10^6$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5$\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
<tr>
<td>4</td>
<td>$5.0 \times 10^5$</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.5$\sigma$</td>
<td>$2^{1/6}\sigma$</td>
<td>$2^{1/6}\sigma$</td>
</tr>
</tbody>
</table>

5.2.2 Measures of Shape

The dimensions of pores in poreated vesicles (vesicles with an opening or pore) are quantified as follows. 20 distinct configurations of each vesicle are considered. For each, the coordinates of beads (retrieved from VMD\textsuperscript{86}) are retrieved that form a ring around the ‘mouth’ of the pore. The gyration tensor is calculated for these coordinates. One eigenvalue of the gyration tensor is uniformly small as the selected beads are nearly in plane. The other two eigenvalues, which denote $L_x^2$ and $L_y^2$ report the square of the two principle axes of an ellipsoidal model of the pore. An effective pore area can then be computed as,

$$Pore\ Area = \pi L_x L_y$$  \hspace{1cm} (34)

Average and standard deviations of this quantity are retrieved for each of the 20 separate configurations for each molecule. This process is repeated for globules designed via each of the two thermodynamic routes as highlighted in Table 5.1.
5.3 Design of Poreated and Tubular Vesicles

Can the motifs in Figure 3.3 be combined to yield more complex structures by leveraging *chain connectivity* to encode features over a range of local geometry/dimensionality? Specifically, this would entail the design of larger-scale structures via the combination of macromers:

\[
(A_\alpha B_\beta)_N + (A_\delta B_\gamma)_M + \ldots
\]

where the system could potentially produce a ‘motif multiblock’ copolymer globule, for instance, consisting of blocks \((A_\alpha B_\beta)_N\) and \((A_\delta B_\gamma)_M\) among others. The use sequence within a long chain to engineer in sterically stabilizing halo chains in earlier sections of this work suggests that this possible, wherein solvophobic and solvophilic segments preferentially occupied specific leaflets in a single-chain vesicle.

As a first step in this direction, a ‘motif diblock’ copolymer was simulated in which one side of the chain consists of a repeating sequence of motif I and the other side consists of a repeating sequence of motif II – i.e., chains of structure \((A_\alpha B_\beta)_n\)-b-(A_\gamma B_\delta)_m. The search space of such systems is massive; moreover, far more complex combinations of these motifs in multiple blocks are clearly possible, involving even larger search spaces. Efficient design in this space may require formal design and machine learning methods such as those recently employed to design sequence-specific copolymer compatibilizers.

Here, simulations of one of these systems: a combination of a worm with a vesicle are investigated. The chain has a structure. \((A_\delta B_2)_{350}\)-b-(A_\delta B_4)_{150}, where the underlying vesicle and worm motifs are A_\delta B_2 and A_\delta B_4 respectively.
As shown by Figure 5.1, this motif diblock accesses a shape not normally accessible via small molecule assembly: a poreated vesicle. This structure results from the worm block lying in the plane of the vesicle surface to stabilize a pore.

For the blocks employed here, simulations indicate an equilibrium between opened and closed pores; over multiple long simulations (approximately $7.5 \times 10^6$ LJ time units or 1.5 billion timesteps per simulation), the molecule spends approximately half its time in each state, with stochastic switching between the two. The switching appears analogous to behavior near the critical micelle condition of lipids: the wormlike component of the chain exhibits coexistence...
between a condensed state in which it forms a pore (an in-plane ‘micelle’) and a disseminated state in which it distributes itself across the vesicle surface, closing the pore. This is illustrated in the bar graph in Figure 5.2.

![Bar graph highlighting pore duration times for thermodynamic routes I (left) and II (right).](image)

Figure 5.2. Bar chart highlighting pore duration times for thermodynamic routes I (left) and II (right).

What other shapes are possible within the design space of this motif diblock copolymer structure, that is, for an arbitrary chain structure: \((A_6B_2)_n-b-(A_4B_4)_m\)? To probe the possibilities in a tractable manner, the vesicle motif is kept constant, \((A_6B_2)_{350}\) and the size of the worm motif, \((A_4B_4)_m\) is varied with \(m=\{100,125,150,175,200,225\}\) as depicted in Figure 5.3. As shown by Figure 5.3, this motif diblock allows access to a multitude of non-trivial shapes: ranging from closed/hollow vesicles to transiently poreated vesicle (as highlighted earlier), poreated vesicle and finally a tubular-like vesicle. These structures bear resemblance to the perforated vesicles or stomatosomes seen in several other works\(^6,103,104\).
The physics of the transition of hollow to open and tubular shapes is as follows. When the fraction of chain coding for worms is small, the worm-coding block distributes across the whole surface of the vesicle; with increasing worm-block concentration, the molecule reaches an intramolecular CMC at which an in-plane worm-coding micelle (the pore) exists in equilibrium with a closed vesicle in which the worm-block is homogenously distributed; for high worm-block concentrations the pore becomes stable. At still higher concentrations (m = 200 and 225), an additional pore forms, yielding a short tubule.

Figure 5.3. Motif diblock copolymer globules, \((A_6B_2)_{350}-b-(A_4B_4)_m\), comprised of blocks individually coding for a vesicle and a worm, respectively. The length of the worm block is varied with form, \((A_6B_2)_m\) with \(m = \{100,125,150,175,200,225\}\). A (solvophobic) beads in motifs I and II are shown as grey and blue, respectively. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries. Each globular conformation is rendered in VMD with front (left) and cutaway side (right) views.
Moving forward, one might ask, can the pore size of poreated vesicles be tuned by controlling the sequence of the motif diblock copolymer globule? Based on the above explanation on the pore opening, it is evident that worm micelles often curl over into toroids, and can exist in plane to produce a pore. From Figure 3.3, it was observed that thinner worms tend to form toroids with smaller ‘donut holes’ – an intuitive consequence of the smaller bending radius of a thinner worm. Based on these reasonings, it is hypothesized that thinner worm-coding blocks might yield smaller pores. As seen in Figure 5.4, two different worm motifs, A3B3 and A5B5, produce tubular and largely poreated respective shapes.

Upon closer inspection of the latter globule in VMD86, the resulting shape resembles more of a ‘bowl’ in most configurations, wherein the ‘pore’ is significantly larger than the solvophobic cavity of the vesicle. On the other hand, the globule with the worm motif, A3B3, produces a tubule as seen in similar structures from Figure 5.3. It should be emphasized that this worm block produces a tubule at a lower DOP (1200 compared to 1400) than the poreated vesicle, (A6B2)350-b-(A4B4)175.

This suggests that this system is above the intramolecular critical micelle concentration to produce an in-plane toroid, and that lowering the DOP of the worm-coding block should yield a poreated vesicle. This turns out to be the case for a worm-coding block, (A3B3)133 and thus an overall chain structure: (A6B2)350-b-(A3B3)133. Comparison of snapshots in Figure 5.5a, and Figure 5.5b indeed suggests that inclusion of the motif coding for a thicker worm indeed leads to a larger pore.
Beyond this reasoning and qualitative analysis, the pore sizes of these two vesicles were quantitatively analyzed as shown in Figure 5.6. Here, the effective pore areas are computed as described earlier in this section’s Methodology. The vesicle with the thicker worm produces a larger pore – approximately 75% larger in both thermodynamic routes. This finding is consistent with the hypothesis – the motif that codes for a thicker worm favors a larger radius of curvature and thus tends to stabilize a larger pore when it forms an ‘in plane toroid’ through an effective intramolecular critical micellization event.

Figure 5.4. Motif diblock copolymer globules, \((A_6B_2)_{350}\)-\(-(A_3B_3)_{200}\), comprised of blocks individually coding for a vesicle and a worm, respectively. Here, the vesicle-coding block is fixed, \((A_6B_2)_{350}\) and three different worm motifs are utilized: \((A_3B_3)_{200}\) (top), \((A_4B_4)_{175}\) and \((A_5B_5)_{280}\). A (solvophobic) beads in motifs I and II are shown as great and blue, respective. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries.
Figure 5.5. Simulation images rendered in VMD of poreated vesicles with sequences (a) \((A_6B_2)_{350}-b-(A_4B_4)_{175}\), and (b) \((A_6B_2)_{350}-b-(A_3B_3)_{133}\). Solvophobic beads are only shown to illustrate the pore, where beads in motifs I and II are shown as grey and blue, respectively.

Figure 5.6. Bar plot highlights the areas for the two poreated vesicles over the two thermodynamic routes: route I (colored red) wherein collapse precedes AB segregation and route II (blue) wherein collapse and segregation occurs simultaneously.
5.4 Large Worm Limit

For a given motif diblock copolymer consisting of vesicle and worm coding motifs, hollow, poreated and tubular shapes can be accessed based on the intramolecular micelle concentration, that is, the overall DOP of the worm coding block. What shapes are possible in the large DOP limit of the worm coding block?

Figure 5.7 shows various motif diblock copolymer globules, at the large worm DOP limit of 2800 for worms: A₄B₄, A₅B₅, and A₆B₄ combined with vesicles: A₃B₂, A₆B₂, and A₈B₂. Of particular importance, the chain structure, (A₆B₂)₃₅₀-b-(A₄B₄)₃₅₀ based on the initial studies in Figure 5.1 and Figure 5.2, produces a large sheet-like structure. The worm-coding block distributes
around the edges of the sheet-coding block and prevents it from curling into a vesicle. For structures based on the thinnest vesicle motif, $A_3B_2$, the vesicle is poreated with multiple holes as seen in earlier sections of this work.

These large-scale structures were only analyzed visually, as it was hard to ascertain what shapes they could be classified into. Overall, this highlights several challenges in developing quantitative measures of complex shapes, as well as prediction of nanoparticular shape based on (multiple motif) sequence alone.

5.5 Summary and Conclusions

In this part of the work, single chain globules previously designed were used as ‘alphabets’ or building blocks for assembly into larger-scale structures.

By leveraging chain connectivity, worm and vesicle coding motifs were combined to produce several combinations of motif multiblock copolymer globules. At a fixed degree of polymerization of the vesicle coding block, a cascade of structures including transiently open/poreated vesicles, poreated and tubular vesicles can be designed via tuning the degree of polymerization of the worm-coding block. The physics of this transition appears due to the location of the system to the critical intramolecular micelle concentration. Below this concentration, the worm is homogeneously distributed across the surface of the vesicle; above, the worm produces an in-plane toroid and results in the design of a poreated vesicle.

Furthermore, it was shown that the pore size can be controlled by tuning the sequence of the worm coding block. Thicker worms have a larger bending radius and tend to produce poreated vesicles with larger pores.

In the large DOP limit, the worm caps the edges of the vesicle motif, preventing the sheet from curling into a vesicle.
Other combinations of vesicle and worm coding motifs were studied. In many cases, however, shape analysis is difficult beyond invariants of the gyration tensor and visual analysis. This position in the research highlights the need for more robust shape metrics, and an efficient structure-searching strategy to design more sophisticated shapes within this design space.
Chapter 6: Molecular Shape Matching

6.1 Introduction

The previous chapter demonstrated the application of multiple motifs (from design rules in Figure 3.3) and leverage of chain connectivity to produce a variety of non-trivial structures such as poreated and tubular vesicles. In future works, trial and error search is not tractable. However, one can leverage the physical understanding of design rules to build ‘alphabets’ of single chain motifs with an efficient structure-searching strategy to design motif multiblock copolymers that assemble into targeted shapes. More complicated shapes such as tadpoles, vesicles with specified solvophilic patterns on the surface and junctions may warrant a more sophisticated set of shape measures beyond the radius of gyration tensor. How can one develop a shape matching technique relevant to molecular trajectories?

Evolution has spawned the development of complex yet efficient biological systems. From an early age, we are taught basic shapes such as squares, triangles, and circles. As we grow, the nervous system continuously evolves in complexity at an exponential rate. But how exactly do biological systems adapt and learn? How can we read jumbled words, and process them into coherent thoughts? How can we distinguish between different shapes drawn with the fuzziest of details?

Even though such questions are evolving areas of research, several biomimetic principles have been utilized to mimic nature’s ability to recognize shapes. Phones are now able to utilize
face recognition to recognize their primary users. Google has developed a doodle tool based on artificial intelligence (AI) that can learn to recognize structures based on very crude sketches.\textsuperscript{105}

In the realm of molecular shape-matching, Osada et al.,\textsuperscript{5} Ballester and Richards\textsuperscript{6} utilized distributions (with dimensionality reduction) of atomic distances based on multiple reference locations including the center of mass. A general challenge with molecular shape matching involves the need for the right alignment or orientation of a test object onto a reference. Alignment-based techniques utilize some level of superposition between the test and target prior to quantifying or determining the shape matching metric. These techniques are usually more accurate but more computationally expensive. Lower dimensional techniques that are alignment-free allow a higher throughput when screening through a database. Kumar and Zhang (2018)\textsuperscript{106} highlighted a variety of shape-similarity methods in the application to drug discovery – specifically for virtual screening, protein structure determination and analysis of electron microscopy images. For a given query or test molecule screened against a database of reference shapes, they classified shape similarity methods into 1) atomic-distance based methods, 2) Gaussian overlay-based methods, and 3) surface-based methods. The similarity between shapes can be calculated based on various metrics such as the Manhattan and Euclidean distance. Notably, in surface-based methods, molecular surfaces are the key shape fingerprints used for matching. Such surfaces may be based on solvent-accessible surface area\textsuperscript{107} or van der Waals surfaces\textsuperscript{108}. From these molecular surfaces, shape distributions or histograms may be utilized for the shape metric in 3D shape matching.

Overall, a robust molecular shape matching should be invariant to scale (or size), rotation and translation and computationally inexpensive to allow rapid screening through a database of reference shapes or structures.
6.2 Algorithm Design

Motivated by shape matching techniques based on molecular surfaces, a three-dimensional molecular shape matching technique was designed in this work. Here, each (test) molecular and solid (3D model) shape is converted to a voxel map. Voxel maps have been previously implemented for use in tracking molecular motion\textsuperscript{109} and object recognition\textsuperscript{110–112}. A voxel is the three-dimensional version of a pixel. The voxel map or array can be a binary or grey area which specifies whether an object is within a voxel (a binary value of 1) or otherwise (a binary value of 0). The voxelization process moves from xyz coordinates to voxel coordinates. Detailed code in Python is provided in the Appendices.

Given the voxel arrays for the solid target (T) and molecule (M), a measure of shape similarity (S) can be calculated via a cell-by-cell product:

\[
S = \frac{\sum_k T_k \cdot M_k}{\sum_k T_k \cdot \sum_k M_k}
\]  

(36)

S will be averaged over multiple alignments and trajectories/snapshots. Here, normalization by the products of each array scales the shape similarity, S, to a value between 0 and 1.

The 3D solid can be constructed in a custom modeling program and can be exported as a .stl file format. Voxelization of the 3D model is beyond the scope of this work and can be implemented via open-source software. Overall, the voxelization process involves first extracting the (x, y, z) coordinates of the vertices and the normal vectors of all triangles/facets in the 3D model and initializing a voxel grid based on the desired number of voxels in each dimension. Then, the coordinate of each voxel is tested to determine whether the point resides inside or outside a triangle of the triangulated surface.
As mentioned in the Introduction, a robust shape-matching algorithm should be invariant to scale, rotation, and translation. One measure of scale is the radius of gyration and can used to scale the test/query object to the target object, given their three-dimensional coordinates. The query can be translated to the target based on equating the systems’ centers of masses. Finally, the query object can be oriented to the same coordinate basis of the target. For a given basis vector, \( v \) and a matrix, \( X \), corresponding to the coordinates of the shape, the new coordinates in the basis are calculated via \( X^{-1}v \). Overall, the molecular shape matching enforces the following:

\[
R_{g,\text{target}} = R_{g,\text{molecule}} \tag{37}
\]

\[
\bar{x}_{\text{target}} = \bar{x}_{\text{molecule}} \tag{38}
\]

where \( R_g \) and \( \bar{x} \) refer to the radius of gyration and centroid respectively.

For the input of a three-dimensional molecular trajectory and a 3D model target shape, the algorithm implementation is as follows. For each trajectory frame:

1. Scale molecule coordinates such that \( R_{g,\text{target}} = R_{g,\text{molecule}} \)
2. Orient molecule so that its eigenvectors form a basis
3. Orient molecule so that its eigenvectors align with target’s eigenvectors (8 possible alignments)
4. Translate resulting scaled and oriented molecule so that its center of mass aligns with target’s center of mass
5. Voxelize target and molecule to some fixed resolution, i.e, number of voxels in each dimension (x,y,z)
6. Use voxel data to calculate a fitness metric, \( S_i \) which will quantify how close the query shape (sequence) is to the 3D model
7. Calculate fitness of each alignment
8. Take the maximum fitness of all alignments as the output for optimization

9. Repeat the above implementation over N trajectories/frames

### 6.3 Results

To test the implementation of the molecular shape-matching algorithm defined above, cylindrical, and curved cylinder target shapes are utilized. The fitness values for four different query shapes are depicted in Table 6.1 and Table 6.2.

Table 6.1. Fitness values for various molecular shapes at different voxelization resolutions against a target cylindrical shape. Each molecular shape consists of 500 points.

<table>
<thead>
<tr>
<th>Label</th>
<th>Molecular Shape</th>
<th>Fitness at n = 25</th>
<th>Fitness at n = 50</th>
<th>Fitness at n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>Sphere</td>
<td>0.1200</td>
<td>0.1200</td>
<td>0.1200</td>
</tr>
<tr>
<td>(ii)</td>
<td>Cylinder</td>
<td>0.4540</td>
<td>0.4720</td>
<td>0.4640</td>
</tr>
<tr>
<td>(iii)</td>
<td>Torus</td>
<td>0.1560</td>
<td>0.1600</td>
<td>0.1520</td>
</tr>
<tr>
<td>(iv)</td>
<td>Curved cylinder</td>
<td>0.3120</td>
<td>0.3100</td>
<td>0.2860</td>
</tr>
</tbody>
</table>

Table 6.2. Fitness values for various molecular shapes at different voxelization resolutions against a target curved cylindrical shape. Each molecular shape consists of 500 points.

<table>
<thead>
<tr>
<th>Label</th>
<th>Molecular Shape</th>
<th>Fitness at n = 25</th>
<th>Fitness at n = 50</th>
<th>Fitness at n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>Sphere</td>
<td>0.4700</td>
<td>0.4900</td>
<td>0.4880</td>
</tr>
<tr>
<td>(ii)</td>
<td>Cylinder</td>
<td>0.9660</td>
<td>0.9700</td>
<td>0.9820</td>
</tr>
<tr>
<td>(iii)</td>
<td>Torus</td>
<td>0.7300</td>
<td>0.7300</td>
<td>0.7280</td>
</tr>
<tr>
<td>(iv)</td>
<td>Curved cylinder</td>
<td>0.9700</td>
<td>0.9880</td>
<td>0.9960</td>
</tr>
</tbody>
</table>
The voxels are visualized in Figure 6.1 and Figure 6.2. From these tests, the fitness values are not strongly sensitive to the number of voxels (resolution) utilized. The shape-matching algorithm works well in the case of the curved cylinder, wherein the curved cylinder matches the target above a fitness value of 0.97. However, it appears that the shape-matching is sensitive to the point density, that is, the total number of points defining the molecular shape. Future improvements could involve interpolation of these points.

6.4 Implementation

The molecular shape-matching algorithm once optimized can be implemented for use in a genetic algorithm for inverse material design. This would replace the fitness functions based on the radius of gyration. As depicted in Figure 6.3, adopted by Patra and coworkers\textsuperscript{82}, an artificial neural network can learn from the growing census linking sequence and fitness (based on shape similarity) to suggest sequences to the genetic algorithm; and thus accelerating the convergence of the genetic algorithm. In the case of a motif diblock copolymer of type: \((A_\alpha B_\beta)_{n-b}-(A_\gamma B_\delta)_m\), one can fix the total number of beads, \(L = n(\alpha + \beta) + m(\gamma + \delta)\) and define a fraction of motif I, \(\Phi = n(\alpha + \beta)/L\). With five free variables:

1. \(\Phi\), number fraction of motif I
2. \(A\), length of motif I
3. \(\alpha\), block length of solvophobic beads in motif I
4. \(B\), length of motif II
5. \(\gamma\), block length of solvophobic beads in motif II

Each variable can be mapped to a 4-bit binary genome, yielding a 20-bit binary genome defining the design space of motif diblock copolymers.
Figure 6.1. Visual rendering of voxel maps of the target cylindrical shape (red) against (i) sphere, (ii) cylinder, (iii) toroid, and (iv) curved cylinder query molecular (all in blue) shapes.

Figure 6.2. Visual rendering of voxel maps of the target curved cylindrical shape (red) against (i) sphere, (ii) cylinder, (iii) toroid, and (iv) curved cylinder query molecular (all in blue) shapes.
6.5 Summary and Conclusions

Motivated by the need for an efficient method of measuring shape and designing more complex shapes within the space of motif multiblock copolymers, this final part of the work is focused on designing a molecular shaping algorithm.

The molecular shape matching algorithm is based on comparing the voxel maps of target and query shapes. Here, a voxel map is an array that tracks the presence (binary value of 1) or the absence (binary value of 0) within a three-dimensional grid defining the dimensions of the molecule. The fitness function is based on the normalized cell-by-cell product of the target and voxel maps. Pre-processing techniques involve scaling the query or test shape to the target’s radius of gyration, translation to the target’s center of mass and orientation to a coordinate basis from the eigenvectors based on the target shape’s gyration tensor. These techniques ensure that the shape matching algorithm is invariant to scale, translation, and rotation.

Testing of the shape matching algorithm with targeted shapes – cylindrical and curved cylindrical 3D models, indicate that the implementation matches the desired query molecular
shapes. However, robustness of the fitness values appears to depend on the point density of the molecular trajectory. Future work involving interpolation methods might be necessary to rectify this issue. Other shape signatures that are alignment-free could be tested for implementation such as atomic distance-based methods. Shape fingerprints such as the radial distribution could also be tested as a metric for testing shape similarity between target and query molecule trajectories.
Chapter 7: Overall Summary and Future Work

The primary objective of this work was based on the manipulation of copolymer sequence to design single-chain nanostructures with targeted shapes.

In the first part of the work, the combined efforts of Brownian dynamics simulations and genetic algorithms were utilized to render the design space of sequence-controlled polymers tractable. Full sequence control was not necessary to design anisotropic nanoglobular shapes such as necklaces, worms, and sheets (which tend to curl into vesicles). In fact, within a bead spring polymer model, the primary factors necessary for complex shape formation were found to involve:

1. Sequence control at a repeat unit level (within 2 to 16 Kuhn monomers)
2. High molecular weight (above a degree of polymerization of 500)
3. High solvent quality contrast (high $\chi_{AB}$)

The solvophobic block length plays a significant role in setting the dimensions of nanoparticular shape. In an exemplifying case of vesicles, larger solvophobic blocks yield vesicles with thicker walls and smaller solvophobic cavitation. The effect of bond stiffness warrants further investigation. Based on results in this work, the effective temperature of the system is an important factor to consider in avoiding crystallization. Generally, it was seen that the shapes achieved here are non-trivial and move beyond the simple classifications of vesicles and sponge-like conformation in the high $\chi_{AB}$ results.

To move towards the design of ‘single chain technologies’ that act independently in solution, single chain nanoparticles were designed to be resistant to aggregation at low
concentration. In the case of vesicles, solvophilic chains were incorporated in the underlying sequence in a sequence-specific manner to promote a high density of solvophilic loops on the vesicle’s surface without perturbing the overall shape. Aggregation behavior of these ‘grafted’ vesicles were investigated, and results demonstrated a level of kinetic stabilization within the timescales probed. Further modeling of the kinetics of the aggregation process is necessary. Preliminary results indicate that the aggregation is a multi-step process, including an initial incubation period.

While many of the single nanoparticular shapes are accessible via multimolecular assembly, this work focuses further on leveraging chain connectivity of several motifs to generate a larger-scale shape, akin to the hierarchical structure in proteins. Results of a motif diblock copolymer globule consisting of worm and vesicle coding blocks demonstrate the design of poreated and tubular vesicles. The pore size of the poreated vesicles can be tuned by controlling the sequence of the worm coding block. The physics of this transition is based on a critical intramolecular micelle concentration, wherein the worm forms an in-plane toroid and thus a pore opening of the vesicle. These vesicles can find potential applications as artificial enzymes or as confined reactive sites.

The design space of motif multiblock copolymers to design larger-scale shapes is massive, limiting the effectiveness of trial-and-error search and design methods. Motivated by these challenges, a molecular shape matching algorithm was designed based on generation of voxel maps defining the three-dimensional structure of nanoparticular shapes. Pre-processing techniques were employed to ensure the matching was invariant to scale, translation, and rotation. With additional testing and iterations, the molecular shape matching algorithm can be utilized in a
neural-network biased genetic algorithm to search for copolymer sequences that assemble into complex hierarchical shapes.
References


(13) Diblock Copolymers


(73) Designing DNA-grafted particles that self-assemble into desired crystalline structures using the genetic algorithm http://www.pnas.org/content/110/46/18431 (accessed 2017-12-28).


(105) Statt, N. This Google-powered AI can identify your terrible doodles https://www.theverge.com/2016/11/15/13641876/google-ai-experiments-quick-draw-image-recognition-game (accessed 2017 -12 -28).


Appendix A: Supplementary Figures

1. Cylindrical-toroidal transformations seen for selected repeating sequences with total length of 10 at an overall DOP = 500

Table A.1. Repeating sequences that exhibit cylindrical-toroidal conformations and their respective block length distributions at DOP = 500

<table>
<thead>
<tr>
<th>Repeating Sequence</th>
<th>Block Length Distribution</th>
<th>‘End-to-End’ Block Length Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBAAAAABBA</td>
<td>{1, 2, 4, 2, 1}</td>
<td>{4, 2, 2, 2}</td>
</tr>
<tr>
<td>BAAAAABBAAB</td>
<td>{1, 2, 4, 2, 2}</td>
<td>{4, 2, 2, 2}</td>
</tr>
<tr>
<td>AAABBAABBA</td>
<td>{3, 2, 2, 2, 1}</td>
<td>{4, 2, 2, 2}</td>
</tr>
<tr>
<td>AABBBAAAABB</td>
<td>{2, 2, 2, 2}</td>
<td>{4, 2, 2}</td>
</tr>
<tr>
<td>AABBBBAABBA</td>
<td>{2, 2, 2, 2, 2}</td>
<td>{4, 2, 2}</td>
</tr>
<tr>
<td>AAAABBAABB</td>
<td>{2, 2, 2, 2}</td>
<td>{4, 2, 2}</td>
</tr>
</tbody>
</table>

Figure A.1. Cylindrical-toroidal conformations seen for repeating sequences in the above table. All images were rendered in VMD.
2. Effect of bond stiffness on nanoparticular shapes of motif homopolymers with repeating motif $A_\alpha B_\beta$ and total length, $\alpha + \beta = 10$

Figure A.2. Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 10$ at $\epsilon = 1.5$ with $6 < \alpha < 10$. A cross-sectional view of some conformations is shown on the right. Beads A and B are colored orange and blue respectively.
Effect of stiffness on nanoglobular shapes for selected sequences with repeating motif, $A_\alpha B_\beta$ for $\alpha + \beta = 10$ at $\epsilon = 1.5$ with $3 < \alpha < 7$. A cross-sectional view of some conformations is shown on the right. Beads A and B are colored orange and blue respectively.
3. Multichain assembly at drag force = 0.01 LJ units

Figure A.4. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $7 \leq \alpha \leq 11$. All images were rendered in VMD.
Figure A.5. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $1 \leq \alpha \leq 6$. All images were rendered in VMD.
4. Multimolecular assembly without drag force

Figure A.6. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $15 \leq \alpha \leq 17$. All images were rendered in VMD.
Figure A.7. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $12 \leq \alpha \leq 14$. All images were rendered in VMD.
Figure A.8. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $7 \leq \alpha \leq 11$. All images were rendered in VMD.
Figure A.9. Comparison of conformations of multimolecular assemblies with single chain copolymer globules with total repeat unit length of $\alpha + \beta = 18$ for $1 \leq \alpha \leq 6$. All images were rendered in VMD.
5. Aggregation Behavior of ‘Grafted’ Vesicles

Table A.2. Exponential fitting parameters for vesicle with repeating motif (i) \( A_6B_2 \), and \([(A_6B_2)_x A_9(B_2)]_{n-1}(A_6B_2)_x A_9 \) for (ii) \( x = 25, y = 50 \), and (iii) \( x = 25, y = 75 \), and (iv) \( x = 10, y = 20 \).

<table>
<thead>
<tr>
<th>Label</th>
<th>( N )</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>25596.74</td>
<td>0.62</td>
</tr>
<tr>
<td>(ii)</td>
<td>86605.17</td>
<td>5.06</td>
</tr>
<tr>
<td>(iii)</td>
<td>121941.8</td>
<td>9.03</td>
</tr>
<tr>
<td>(iv)</td>
<td>170404.4</td>
<td>13.07</td>
</tr>
</tbody>
</table>
6. Motif Diblock Copolymer Globules

Table A.3. Motif diblock copolymer globules simulated with total DOP \( \approx 5600 \)

<table>
<thead>
<tr>
<th>Worm Block</th>
<th>Vesicle Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>((A_4B_4)_{350})</td>
<td>((A_3B_2)_{560})</td>
</tr>
<tr>
<td>((A_4B_4)_{350})</td>
<td>((A_6B_2)_{350})</td>
</tr>
<tr>
<td>((A_5B_5)_{280})</td>
<td>((A_3B_2)_{560})</td>
</tr>
<tr>
<td>((A_5B_5)_{280})</td>
<td>((A_6B_2)_{350})</td>
</tr>
<tr>
<td>((A_6B_6)_{233})</td>
<td>((A_3B_2)_{560})</td>
</tr>
<tr>
<td>((A_6B_6)_{233})</td>
<td>((A_6B_2)_{350})</td>
</tr>
<tr>
<td>((A_6B_4)_{280})</td>
<td>((A_6B_2)_{350})</td>
</tr>
<tr>
<td>((A_6B_4)_{280})</td>
<td>((A_8B_2)_{280})</td>
</tr>
<tr>
<td>((A_6B_4)_{280})</td>
<td>((A_8B_2)_{280})</td>
</tr>
<tr>
<td>((A_6B_4)_{280})</td>
<td>((A_8B_2)_{280})</td>
</tr>
</tbody>
</table>
Figure A.10. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, $(A_4B_4)_{350}$. A (solvophobic) beads in motifs I and II are shown as great and blue, respective. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries.
Figure A.11. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, (A₅B₅)₂₈₀. A (solvophobic) beads in motifs I and II are shown as grey and blue, respectively. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries.
Figure A.12. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, \((A_6B_6)_{233}\). A (solvophobic) beads in motifs I and II are shown as great and blue, respective. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries.
Figure A.13. Various motif diblock copolymer globules, comprised of blocks individually coding for a vesicle and with fixed worm block, \((A_6B_4)_{280}\). A (solvophobic) beads in motifs I and II are shown as great and blue, respective. B (solvophilic) beads in motifs I and II are rendered smaller for clarity and are shown as green and ice blue, respectively. A beads in motifs I and II (grey and blue beads) are chemically identical, as are B beads in motifs I and II (green and ice blue beads) – assembly of the two motifs is driven purely by sequence and not by use of distinct chemistries.