JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau

Article

Computational Reverse Engineering Analysis for Scattering Experiments (CREASE) on Vesicles Assembled from Amphiphilic Macromolecular Solutions

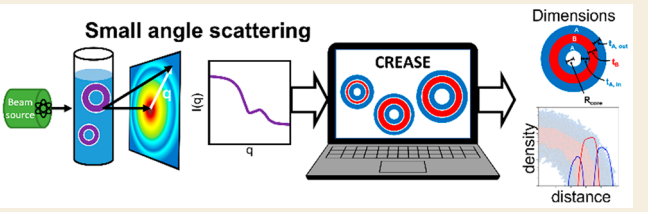
Ziyu Ye, Zijie Wu, and Arthi Jayaraman*

Cite This: *JACS Au* 2021, 1, 1925−1936

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this paper we present the development and validation of the "Computational Reverse-Engineering Analysis for Scattering Experiments" (CREASE) method for analyzing scattering results from vesicle structures that are commonly found upon assembly of synthetic, biomimetic, or bioderived amphiphilic copolymers in solution. The two-step CREASE method takes the amphiphilic polymer chemistry and small-angle scattering intensity profile, $I_{exp}(q)$, as input and determines the vesicles' structural features on multiple length scales ranging from assembled vesicle wall's individual layer thicknesses to the monomer-level packing and distribution of polymer conformations. In the first step of CREASE, a genetic algorithm (GA) is used to determine the relevant vesicle dimensions from the input macromolecular solution information and $I_{exp}(q)$ by identifying the structure whose computed scattering profile best matches the input $I_{exp}(q)$. Then in the second step, the GA-determined dimensions are used for molecular reconstruction of the vesicle structure. To validate CREASE for vesicles, we test CREASE on input scattering intensity profiles generated mathematically (termed as *in silico* $I_{exp}(q)$ vs q) from a variety of vesicle sizes with known dimensions. We also test CREASE on *in silico* $I_{exp}(q)$ vs q generated from vesicles with dispersity in all relevant dimensions, resembling real experiments. After successful validation of CREASE, we compare the CREASE-determined dimensions against those obtained from the traditional approach of fitting the scattering intensity profile to relevant analytical model in SASVIEW package. We show that CREASE performs better than or as well as the core−multishell analytical model's fitting in SASVIEW in determining vesicle dimensions with dispersity. We also show that CREASE provides structural information beyond those possible from traditional scattering analysis using the core−multishell model, such as the distribution of solvophilic monomers between the vesicle wall's inner and outer layers in the vesicle wall and the chain-level packing within each vesicle layer.



KEYWORDS: *small angle scattering, vesicle, computational analysis, CREASE, amphiphilic polymer*

I. INTRODUCTION

Vesicles are spherical nanostructures with a characteristic hollow-core−shell architecture enclosed by a wall membrane. Owing to their hollow structure, vesicles make up an important class of self-assembled nanostructures capable of encapsulating materials that enable their use in a wide range of applications such as nanoreactions and drug delivery.[1−6] In particular, the self-assembly of vesicles from amphiphilic molecules like surfactants, synthetic block copolymers, and bioinspired macromolecules (e.g., peptide−lipid and peptide−polymer conjugates)[4,7−14] pave the way for designing customizable, smart nanocarriers with controlled functions. To facilitate fundamental understanding for improved engineering of vesicles from novel polymeric materials, accurate characterization of the assembled structure, from the macromolecular to the microscopic length scale, is critical.

Experimental imaging techniques for assembled macromolecular structures include transmission electron microscopy (TEM),[15] cryogenic TEM,[16−18] and atomic force microscopy (AFM).[19] With these microscopy techniques, the real space shapes and distances are easily accessible, but the images in most cases are limited to a two-dimensional slice of the assembled structure and small length scales are often difficult to resolve. In contrast, small-angle scattering (SAS) techniques like neutron (SANS) and X-ray (SAXS) scattering can resolve structural features over a wide range of length scales from micrometers down to a few nanometers. In particular, SAS techniques are powerful tools for polymers and protein assemblies in solutions due to the length scales relevant to these structures and the noncrystalline nature of these materials under common conditions.[17,18,20−32] These scattering experiments produce the scattering intensity $I(q)$, as a

Received: July 8, 2021
Published: September 29, 2021

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

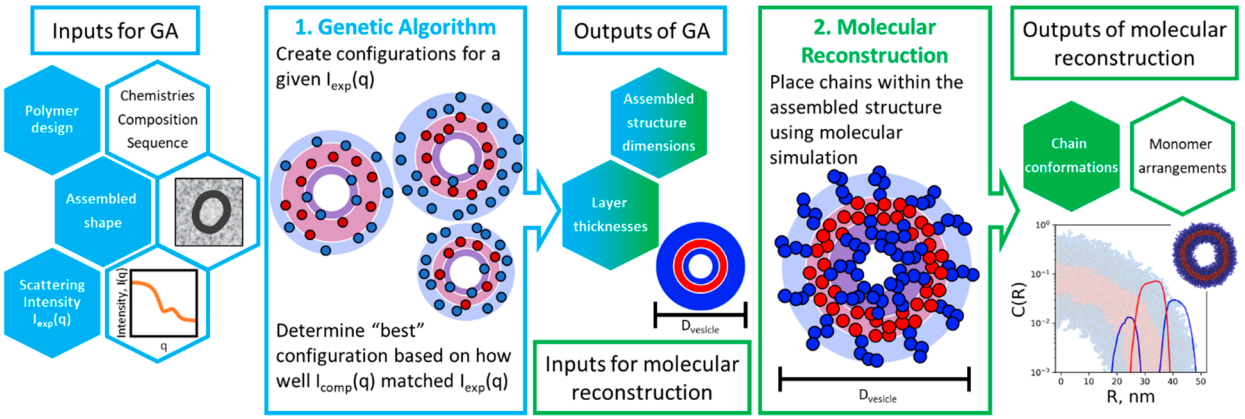


Figure 1. Schematic flowchart of the steps in the Computational Reverse Engineering Analysis for Scattering Experiments (CREASE).

function of wavevector q, from the collision of X-rays[33] or neutrons[34−37] with the sample. Traditionally, the potential of using SAS techniques to determine structural features is locked behind the interpretation of $I(q)$ vs q by fitting the $I(q)$ vs q data with relevant analytical models. While the existing analytical models perform well for scattering profiles from canonical assembled structures and conventional polymer chemistries, they may not work in getting important dimensions of the structure for novel polymer chemistries and architectures and/or previously unseen assembled structures. The molecular level information, like the distribution of polymer chain conformations, also might not be accessible to models that tend to determine average chain conformations. Lastly, given the advances in high-throughput instrumentation for fast characterization of many samples, there is also a need for alternate computational methods that do not rely heavily on manual fitting of the scattering data with analytical models. To address these existing limitations in the analysis of SAS results, we recently developed the CREASE approach to analyze scattering results without relying on approximate analytical models for fittings to the scattering profile. This approach has already been successfully applied to analyze the scattering results for dilute amphiphilic polymer solutions with spherical, cylindrical, fibrillar, and elliptical cylinder micelles.[38−40] In this paper we extend this CREASE method to analyze scattering profiles from vesicle structures and compare the results obtained from CREASE to those from traditional fitting approaches to demonstrate that CREASE works well and provides information about vesicle structure beyond those obtainable from traditional approaches involving fitting with existing analytical models.

CREASE is a two-step computational method that combines a genetic algorithm (GA) in step one with molecular simulations in step two to capture structural features from the assembled aggregate to the chain- and monomer-level packing. GA is an optimization method inspired by the natural genetic evolution process where "fit" individuals in a generation continue to exist in future generations and "unfit" individuals do not, and eventually the best individuals survive. GA has been used in science and engineering to generate the "best" or "optimum" solutions to complex optimization problems. In the context of CREASE, the first step involves a GA to identify the "best" structure whose computed scattering profile matches the input experimental scattering profile. In this GA step, we take as input the polymer chemistries, an approximate shape of the structure (e.g., from imaging techniques), and the experimental scattering profile, $I_{exp}(q)$, and generate a series of potential vesicle configurations to optimize toward the "best" vesicle configuration and its dimensions (e.g., diameter of the vesicle and the layers' thicknesses). Then in the second step, we reconstruct the assembled structure using molecular simulations with either coarse-grained or atomistic models of polymer chains of input chemistry within the GA determined dimensions to understand the microscopic details of the vesicle structure. We extract the relative distribution of monomers in various domains, the distribution of chain configurations, and molecular packing information, which are not accessible when fitting with analytical models.

In this study, we extend this CREASE approach to vesicles with a hollow-core and a three-layer shell morphology. We first apply the CREASE approach on *in silico* $I_{exp}(q)$ generated from vesicles with known dimensions to quantitatively test the predictive capabilities of CREASE against those from the traditional approach of fitting using a core−multishell analytical model. We also demonstrate that in addition to capturing the relevant vesicle dimensions (e.g., core radius, thickness of each of the three layers), CREASE determines information important to the vesicle morphology that is not readily available to traditional (e.g., core−multishell) analytical models such as the ratio of molecules distributed in the various layers of the vesicle shell. To demonstrate the potential of applying this CREASE approach to samples prepared under real experimental conditions where dispersity in dimensions can be found, we extend the GA step to include dispersity in various dimensions and test the GA step on *in silico* $I_{exp}(q)$ generated from disperse vesicles with known dimensions. Finally, in the second step of CREASE, we reconstruct at the molecular level a representative slice of the vesicle structure for two cases of amphiphilic molecules and extract the chain- and monomer-level information.

To the readers who have followed the previous publications on CREASE[38,40] we note the following new aspects of CREASE presented in this study that were absent in those previous studies: in the CREASE method extended to vesicles in this paper we demonstrate the ability to (a) characterize the level of dispersity present in relevant dimensions of the assembled structure which was not demonstrated in the previous implementation of CREASE on spherical[38] and cylindrical micelles[40] and (b) obtain additional information from the GA step of CREASE regarding the distribution of the macromolecules between the two solvophilic layers of the

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

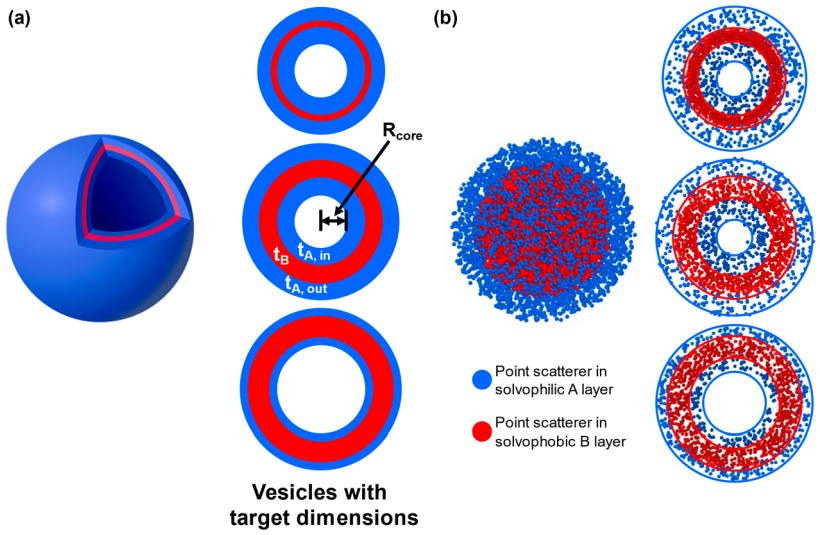


Figure 2. Generation of *in silico* $I_{exp}(q)$ from vesicles with known target dimensions in the solvophilic A (blue) and solvophobic B (red) layer. (a) Relevant dimensions of the vesicle comprised of the inner (t_{Ain}) and outer (t_{Aout}) solvophilic A layers and a middle solvophobic B layer (t_B). (b) Placement of scatterers (dots) within the target dimensions (solid lines) of the solvophilic A and solvophobic B layers visualized for the complete vesicle (left) and a slice from the vesicle center (right). These scatterer positions are used to obtain the *in silico* $I_{exp}(q)$ using eqs 1−4 as described in the text.

vesicle bilayer. This extension of CREASE to vesicles and the additional new aspects of CREASE characterizing the distribution of the molecules within the vesicle layers and determination of the dispersity hold important insights for scientists wishing to link molecular-level structural information of assemblies to macroscopic properties of the assembled structures (e.g., transport of drugs through the layers of the vesicles). To enable broad dissemination and use of this CREASE method, we have now made the code for the GA portion of CREASE open source by sharing on github [https://crease-ga.readthedocs.io] a step-by-step tutorial and the entire code of the GA portion of the CREASE method as applied to vesicles. In that tutorial, we also guide users who may wish to adapt this CREASE-GA code to other assembled shapes.

II. APPROACH

II.A. CREASE Method Overview

We present here the two steps of CREASE and provide details of the key extensions made beyond the previous CREASE work[38−40] for interpreting the scattering intensity profiles of vesicle structures that assemble from amphiphilic block copolymers (BCPs) or peptide−peptide conjugates in solution. This two-step computational method, as summarized in the flowchart shown in Figure 1, interprets a small angle scattering profile by determining the structural features on multiple length scales from the assembled structure dimensions of vesicle size and individual layer thickness within the vesicle wall to the monomer packing and distribution of chain conformations. As input, CREASE takes in the molecule design information, the experimental scattering intensity (e.g., SAXS, SANS), and the assembled structure shape, which can be obtained from real-space imaging methods (e.g., TEM, cryo-TEM). The first step of CREASE uses a genetic algorithm (GA) to determine the relevant macroscopic vesicle dimensions from the input information. Then in the second step, we use the GA-determined dimensions to guide the molecular reconstruction of the vesicle assembly by molecular dynamics (MD) simulations to capture the chain conformations of the polymer or peptide chains within the assembly. In this work, we apply CREASE to the scattering intensity profile of spherical vesicles that have a hollow core enclosed by one wall containing three distinct layers where a solvophobic B layer is sandwiched between two solvophilic A layers.

The input to CREASE consists of the BCP or peptide−peptide conjugate design, the experimental scattering intensity profile, $I_{exp}(q)$, and an approximate shape of the assembled structure obtained using another imaging technique besides scattering. The design information of the macromolecule (BCP or peptide) with solvophilic A and solvophobic B blocks includes the chemistries, sequence (e.g., AB, ABA), molar masses (M_A, M_B), repeat unit's contour length ($l_{mono,A}$, $l_{mono,B}$), chain composition or fraction of A repeat units in the chain (f_A), and the total number of repeat units (N). Some of these inputs are used in the GA step to determine the computed scattering intensity, $I_{comp}(q)$, of the individual structure that best matches the input $I_{exp}(q)$. Then in the second step, the macromolecular design information (e.g., N, sequence, f_A, A and B chemistry, molar masses, contour lengths) guides the molecular reconstruction step. For this study, we use *in silico* $I_{exp}(q)$ vs q of vesicles as the input scattering intensity profile to facilitate testing and quantitative validation of each individual step of CREASE. In real implementation, one would use experimentally obtained $I_{exp}(q)$ vs q, as demonstrated in earlier implementation of CREASE on spherical and fibrillar micelles.[38,39] The last piece of input to CREASE is the approximate shape of the assembled structure, which can be specified using measurements from real-space characterization techniques (e.g., microscopy).

II.A.1. Step 1: Genetic Algorithm (GA). The GA step is used to determine the macroscopic dimensions of the vesicle that will produce a computed scattering intensity profile, $I_{comp}(q)$, that best matches the $I_{exp}(q)$. We initiate the algorithm by creating a population with many candidate vesicle configurations (individuals) in the population. Each individual's genetic information includes relevant vesicle

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

dimensions—core radius (R_{core}), inner A layer thickness (t_{Ain}), B layer thickness (t_B), and outer A layer thickness (t_{Aout}). In the GA step, the vesicle configurations do not hold or optimize any A−B chain-level information, and the relevant structural dimensions are obtained purely using scatterers (as described in the next paragraph and shown in Figure 2). Besides the vesicle dimensions, the split of the A scatterers occupying the inner (s_{Ain}) and outer solvophilic layers of the vesicle's wall, which corresponds to the distribution of the solvophilic monomers in the inner and outer layers, is also optimized during the GA search. Additionally, we include $I_{background}$ for each individual to determine the experimental background scattering intensity. Initially, the values for each of these parameters holding the genetic informations are randomly assigned for each individual from a set range of values for each parameter that is selected based on the input information about the macromolecules. We direct the reader to the Supporting Information (SI) Section I.A, where some of the details of the implementation of the genetic information in the GA are provided.

Each generation of the GA consists of a certain number of individuals (i.e., the population size) which is a choice that the user must make to balance between choosing a large number of individuals per generation to improve the genetic diversity in the population and choosing a smaller number of individuals per generation to keep it computationally feasible. For our system, we find that using 80 individuals per generation provides enough diversity in the population to arrive at a consistent optimal answer across 10 independent GA trials for a given input $I_{exp}(q)$ and correctly reproduces the target dimensions within reasonable computational time.

Next, we calculate the $I_{comp}(q)$ for each individual in the population by randomly placing A and B type scatterers within its vesicle layer boundaries. The ratio of A and B scatterers matches the composition of the amphiphilic macromolecule (f_A) given as an input to CREASE. These A and B scatterers could be point scatterers or scatterers with a specified diameter. Regardless of the choice, the total number of A and B scatterers is another choice the user must make to balance the accuracy of the output vesicle dimensions from GA and the computational intensity of the $I_{comp}(q)$ calculation which involves computations over all pairs of scatterers. Typically for the vesicles presented in this study, the total number of scatterers ranges from 500 to 5000 per individual and is highly variable due to the range of different individual vesicle candidate sizes. Using the placements of the N_A number of solvophilic A and N_B number of solvophobic B scatterers, the intravesicle structure factor, $\omega(q)$, is calculated as

$$\omega(q) = \left\langle \frac{1}{N_A + N_B} \sum_{i=1}^{N_A+N_B} \sum_{j=1}^{N_A+N_B} \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle \tag{1}$$

where r_{ij} is the distance between the pair of scatterers i and j. For each individual, we calculate its $\omega(q)$ on seven different configurations of random scatterer placements and use the averaged $\omega(q)$ to produce its $I_{comp}(q)$. Having multiple configurations of scatterer placements mitigates the chance of any one configuration of scatterer placements within the vesicle dimension biasing the computed scattering intensity. One could also choose to replace the use of multiple (seven in our case) placements of fewer scatterers with just one placement using a significantly higher total number of scatterers per unit volume (i.e., scattering density); the latter case may reduce the biases one may introduce with placements of few scatterers or at smaller scatterer density. However, at higher number of scatterers, the associated computational time for performing the $\omega(q)$ calculation also rises significantly. Another reason we choose to use multiple (seven) different configurations of random scatterer placements is to incorporate the ability to capture dispersity in vesicle dimensions as described in later paragraphs.

Using the averaged intravesicle structure factor $\omega(q)$, we then calculate the form factor as

$$F_M^2(q) = \sum_{\alpha\in\{A,B\}} \sum_{\beta\in\{A,B\}} b_\alpha b_\beta F_{GA,\alpha}(q) F_{GA,\beta}(q) \omega(q) \tag{2}$$

where b_A and b_B are the scattering length densities of A and B scatterers, respectively. The values of b_A and b_B are set to 1 in our code as we are using hypothetical scatterers. When using scatterers with a defined diameter, the sphere form factor amplitudes $F_{GA,A}$ and $F_{GA,B}$ for the A and B scatterers are given by

$$F_{GA,\alpha}(q) = \frac{3[\sin(q0.5\sigma_{GA,\alpha}) - q0.5\sigma_{GA,\alpha}\cos(q0.5\sigma_{GA,\alpha})]}{(q0.5\sigma_{GA,\alpha})^3} \tag{3}$$

where α is A or B and $\sigma_{GA,\alpha}$ is the corresponding scatterer diameter. When using point scatterers (i.e., infinitely small diameter), $F_{GA,\alpha}(q) = 1$.

Using $F_M^2(q)$, we then calculate the computed scattering intensity, $I_{comp}(q)$, as

$$I_{comp}(q) = F_M^2(q) S_{MM}(q) + I_{background} \tag{4}$$

where $S_{MM}(q)$ is the intervesicle structure factor and $I_{background}$ is the background scattering intensity. As the vesicle assemblies of interest to most applications are in dilute solutions with negligible interaction between vesicles, we use $S_{MM}(q) = 1$. The range of q values (0.003−0.1 Å^{-1}) that we use for the calculation of $I_{comp}(q)$ spans the diameter of the whole vesicle on the low q side to the thinnest layer dimension within the wall of the vesicle on the high q side. Besides choosing the range of q values via guidance from the likely vesicle dimension based on real-space measurements of the vesicle from methods like microscopy, it is best to test the sensitivity of the GA results to the range of q values used to calculate $I_{comp}(q)$ for comparison with $I_{exp}(q)$.

Having calculated the $I_{comp}(q)$ for each individual in the population, the GA determines how well a particular individual's $I_{comp}(q)$ matches the input $I_{exp}(q)$ through the fitness of that individual. A quantitative measure of the difference between $I_{comp}(q)$ and $I_{exp}(q)$ can be calculated through the weighted sum of the squared log difference between $I_{exp}(q)$ and $I_{comp}(q)$ as given by

$$\text{sse} = \sum w \left[\log\left(\frac{I_{exp}(q)}{I_{comp}(q)} \right) \right]^2 \tag{5}$$

where the weighting factor w is defined as $w = \log(|q_{i+1} - q_i|)$ which assigns weights to $I(q)$ considering the log-scale spacing of the q values. We then obtain the fitness function as

$$\text{Fitness} = X(\text{sse}_{max} - \text{sse}) + Y \tag{6a}$$

where sse_{max} is the maximum value of the sse for an individual in the generation and X and Y are scaling factors that prevents

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

low fitness solutions from being eliminated prematurely in the GA. X and Y take on the expressions

$$X = (\mathrm{cs} - 1)\frac{(\mathrm{sse}_{\max} - \mathrm{sse})}{\max(\mathrm{sse}_{\max} - \mathrm{sse}) - \mathrm{average}(\mathrm{sse}_{\max} - \mathrm{sse})} \tag{6b}$$

and

$$Y = (1 - x)\mathrm{average}(\mathrm{sse}_{\max} - \mathrm{sse}) \tag{6c}$$

where cs = 10 is a scaling constant.

We note here that the user may choose to use a different form of error as long as the choice of error function ensures appropriate error weighting in the range of q explored. For studies where the target $I_{\mathrm{exp}}(q)$ comes from experimental scattering with an associated uncertainty at each q, a commonly used formulation is χ^2 to account for the intrinsic uncertainties in $I_{\mathrm{exp}}(q)$ during fitting optimization. For example, in SASVIEW one uses χ^2 to quantify the goodness of fit by taking the direct difference squared between the computed and target data set over the fitted q range weighted by the squared error in $I_{\mathrm{exp}}(q)$ at each q.[41] If the error associated with each q is known, χ^2 can similarly be implemented into the fitness function in CREASE to account for uncertainty in $I(q)$ at each q.

Based on the calculated fitness, the GA selects the individuals in the population that get to move on to the next generation in the GA. The best individual, identified as having the highest fitness value in the current generation, always gets selected for the next generation. The rest of the individuals are selected with a probability proportional to the fitness of the individual. The selected individuals additionally have a chance to undergo crossover and/or mutation. With crossover, two desirable parent individuals' features are combined to produce an offspring that can improve the solution set. Mutation brings diversity to the population by changing random features of the individuals. We direct the reader to the SI Section I.B where the details of the implementation of crossover and mutation in the GA are provided. The GA continues the selection process until the best individual's fitness and the average population fitness have converged over multiple generations. The relevant dimensions (R_{core}, t_{Ain}, t_{B}, and t_{Aout}) of the individual with the highest fitness in the last generation are reported for the 10 independent GA trials.

II.A.1.1. Incorporation of Dispersity in Vesicle Dimensions in the GA Step. Dispersity in dimensions is commonly seen in experiments of assembled vesicles. Hence, we extend CREASE to account for dispersity by incorporating a feature to handle multiple contributions to the vesicle core dimension into the GA during the search for the best individual. In the calculation of $\omega(q)$ for a given individual with core radius R_{core}, each of its seven different configurations of random scatterer placements is created using a core radius sampled from a distribution of R_{core} values centered about the averaged core radius value with a standard deviation of σ_{core}. The averaged $\omega(q)$ used for obtaining the computed scattering intensity for an individual thus incorporates dispersity in the vesicle core dimensions. The GA searches for the value of the σ_{core} parameter along with the different vesicle dimensions when determining the best individual. In the results section we show results from two versions of CREASE, one that does not account for dispersity and one that incorporates dispersity during the GA step, on *in silico* $I_{\mathrm{exp}}(q)$ created with dispersity in the target dimension(s).

II.A.2. Step 2: Molecular Reconstruction. Having determined the relevant vesicle dimensions from the GA in step 1, we use molecular dynamics (MD) simulations to reconstruct the vesicle assembly. To retain the important features of a vesicle enclosed by a wall membrane while maintaining computational feasibility, we choose to model the vesicle wall as a slice of the vesicle from its equator instead of simulating the entire vesicle. This gives a cylindrical slice geometry that retains the same core radius, individual layer thickness, and curvature in the slicing plane direction as the whole vesicle itself (Figure S1). For the reader to appreciate the computational speed of the MD simulations of a narrow slice of the vesicle containing 194 400 particles, we note that it requires 140 h to run 10 000 000 steps (including equilibration of chains in the vesicle to collapsed conformation, collection of configurations for analysis) using 1 compute node (with 20 cores) on the University of Delaware Farber supercomputing cluster that contains hundreds of compute nodes with dual Intel E5-2670v2 (Ivy Bridge) processors, 20 cores per node, 64 GB RAM (1866 MHz), an FDR InfiniBand network fabric, and a 256TB Lustre filesystem. While the vesicle slice can maintain computational efficiency compared to simulating an entire vesicle and sufficiently capture the most important features of the complete vesicle, an infinitely long cylindrical geometry may give rise to differences in amphiphile conformations as compared to a cylindrical shell. In cases where it would be feasible to simulate the entire sphere (*e.g.*, small vesicle systems), one could perform the reconstruction step to capture the complete vesicle.

In principle, one could use an atomistic or any relevant coarse-grained (CG) model for the reconstruction, but in this study, to demonstrate the types of information one can obtain from molecular reconstruction in CREASE, we use CG models of two types of amphiphilic polymers (Figure S1): (1) amphiphilic, linear BCP, A_{18}-*b*-B_{36}, modeled with a flexible CG bead−spring model[42] with A being the solvophilic chemistry and B the solvophobic chemistry and (2) solvophobic (B) elastin-like polypeptides (ELP) conjugated to solvophilic (A) collagen-like polypeptides (CLP) represented with the CG model of Condon et al. that captures the stiff CLP triple helix as a rigid rod and three ELP strands each as a flexible, bead−spring chain.[43]

The bonded interaction between beads on either CG polymer chain (A or B block of the linear diblock or ELP chain) is governed by the harmonic spring potential $U_{\mathrm{bond}}(r) = k_{\mathrm{bond}}(r - r_{ij}^0)^2$ with an equilibrium bond length $r_{ij}^0 = 1.0\sigma$, where σ is the bead diameter, and a bond stretching force constant $k_{\mathrm{bond}} = 500kT/\sigma^2$, where k is the Boltzmann constant and T is the simulation temperature in reduced Lennard-Jones (LJ) units.[44] For the ELP-CLP model, we fix the beads in the CLP chains at a constant bond length of 1.0σ with a rigid-body integrator.[45,46] We model the solvent implicitly in our simulations using the pairwise, nonbonded interaction between the beads to capture the solvophilic and solvophobic nature of the CG chains. For the solvophobic beads (B or ELP) in a poor implicit solvent, the B−B interactions are modeled with an attractive Lennard-Jones (LJ) potential with a cutoff of 2.5σ. To capture the effect of an implicit good solvent for the solvophilic beads (A or CLP), we model the A−A and A−B interactions using the purely repulsive Weeks−Chandler−Andersen (WCA) potential with a cutoff of $2^{1/6}\sigma$.[47] We capture the solvophobicity of the B (or ELP) chains in the assembled state of the vesicle by setting the B−B interaction

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

strength (ε_{BB}) to 0.8ε. We perform CG MD simulations using the LAMMPS package in the *NVT* ensemble with the Nosé–Hoover thermostat at T = 1.[45] After placing the linear BCP chains or the ELP-CLP trimers within the cylindrical slice geometry shown in Figure S1, we relax the system (typically consisting of 100 000 to 200 000 particles) for 1 000 000 timesteps to allow the solvophobic chains to equilibrate to their collapsed conformation state. We then sample configurations every 100 000 timesteps after equilibrium has been reached for determining the distribution of radius of gyration and the monomer concentration profile as a function of the radial distance from the center of the vesicle.

II.B. Generating *In Silico* $I_{exp}(q)$ with Known Dimensions

We create a set of *in silico* $I_{exp}(q)$ for vesicles that have varying dimensions in the core radius (R_{core}), inner A layer thickness (t_{Ain}), B layer thickness (t_B), and outer A layer thickness (t_{Aout}) to test the performance of CREASE in determining the relevant features in a vesicle structure (Figure 2a). We choose these cases of vesicle dimensions to cover a variety of combinations in the relevant features:

1. Small core (R_{core} = 10 nm), thick A layers (t_{Ain} = t_{Aout} = 12 nm), thin B layer (t_B = 6 nm)
2. Small core (R_{core} = 10 nm), thick A layers (t_{Ain} = t_{Aout} = 12 nm), thick B layer (t_B = 12 nm)
3. Medium core (R_{core} = 20 nm), thin A layers (t_{Ain} = t_{Aout} = 6 nm), thick B layer (t_B = 12 nm)
4. Large core (R_{core} = 30 nm), thick A layers (t_{Ain} = t_{Aout} = 12 nm), thin B layer (t_B = 6 nm).

We then generate $I_{exp}(q)$ for each vesicle by placing scatterers within the target dimensions that we have chosen for the vesicle (Figure 2b) and calculating *in silico* $I_{exp}(q)$ using eqs 1–4. To describe the sample under real conditions where there is dispersity in the vesicle size, the *in silico* $I_{exp}(q)$ is created with dispersity in the target dimension of the core radius first assuming the major contribution to the variance in vesicle sizes comes from the core dimension. We take the *in silico* $I_{exp}(q)$ of seven vesicles with varying core sizes and use their averaged dimension as the target vesicle dimension. The dimensions of the seven vesicles are chosen such that the standard deviation in the core radius is 20% of the target value for the core radius. Lastly, using vesicle cases 1–3 where the total vesicle diameters are similar, we also generate a highly disperse sample in which the *in silico* $I_{exp}(q)$ is taken as the averaged $I(q)$ from the three vesicle cases. This creates an input scattering intensity profile where the target value in all vesicle dimensions (R_{core}, t_{Ain}, t_B, t_{Aout}) has an associated uncertainty, with the biggest variance in the R_{core}. We then perform CREASE and compare the result to core–multishell model fit to this set of *in silico* $I_{exp}(q)$.

III. RESULTS AND DISCUSSION

III.A. Analysis of *In Silico* Scattering Profiles for Vesicles with Dispersity in Dimensions

We select three vesicle systems with similar total vesicle diameter and varying dimensions in the core radius, inner and outer solvophilic layers, and middle solvophobic layer to demonstrate CREASE's ability in determining and distinguishing each of the four target dimensions (R_{core}, t_{Ain}, t_B, and t_{Aout}). For each of the three vesicle cases of (1) small core with thick A layers and thin B layer (Figure 3a,b), (2) small core with thick A and B layers (Figure 3c,d), and (3) medium core with thin A layers and thick B layer (Figure 3e,f), we use the GA component of CREASE to determine the vesicle dimensions from the input *in silico* $I_{exp}(q)$ generated from vesicles with 20% dispersity in the core radius (R_{core}) dimensions. This captures the kind of dispersity seen in experimental samples of vesicle assemblies from solutions or amphiphilic polymer or peptide materials.[9,11] First, we observe how well the computed scattering intensity $I_{comp}(q)$ of the best individual from successive generations in the GA matches the $I_{exp}(q)$ in the range of q values that span from a lower bound of 0.003 Å^{-1}, corresponding to a real-space length scale greater than the vesicle diameter of the target assembly, to the higher bound of 0.1 Å^{-1}, corresponding to the real-space length scale of the smallest dimension of interest in the vesicle (i.e., the individual layer thickness). We note once more that the choice of the range of q values may not be obvious when working with real experimental $I_{exp}(q)$, and thus, it is best to compare the GA results for multiple ranges of q values. As shown in Figure

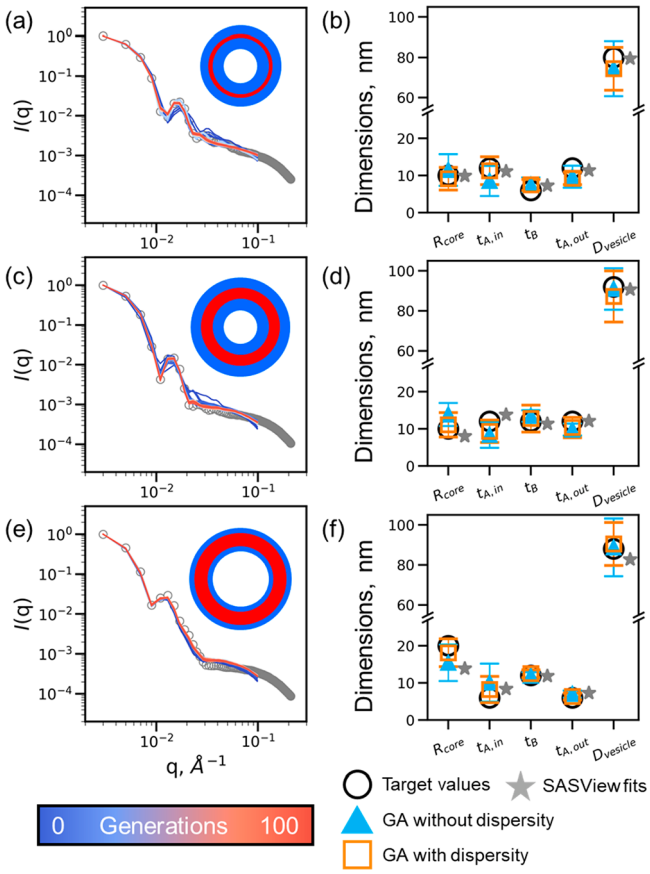


Figure 3. CREASE's analysis on three cases of vesicles shown schematically with the target dimensions: (a, b) small core with thin B layer, (c, d) small core with thick B layer, and (e, f) medium core with thick B layer. (a, c, and e) $I_{comp}(q)$ of the best individuals (colored lines) from all generations in one example GA run (with dispersity) compared with the input *in silico* scattering intensity $I_{exp}(q)$ (open circles). Progression from the initial generation to the final generation is shown with increasingly warmer colors. (b, d, and f) GA-determined dimensions of relevant vesicle features for the cases where dispersity in dimensions is not accounted for in the GA step (blue triangles) and when dispersity is incorporated into the GA (orange squares). The target dimensions of the *in silico* $I_{exp}(q)$ are shown as open circles with 20% dispersity in the core radius (R_{core}) dimensions. SASVIEW fitting results using the core–multishell model (gray stars) are offset to the right of the target dimensions.

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925–1936

Table 1. GA-Determined Parameters Describing the Split of Solvophilic A Chains on the Inside of the Vesicle Wall (s_{Ain}) and Dispersity (σ_{core}) in the Core Radius Dimension[a]

	s_{Ain}			σ_{core} (%)		
vesicle case	target	GA no dispersity	GA with dispersity	target	GA no dispersity	GA with dispersity
small core thin B layer	0.2	0.28 ± 0.04	0.28 ± 0.03	20	N/A	24 ± 3
small core thick B layer	0.2	0.26 ± 0.03	0.34 ± 0.03	20	N/A	18 ± 3
medium core thick B layer	0.2	0.30 ± 0.03	0.24 ± 0.04	20	N/A	24 ± 2
highly disperse	0.2	0.26 ± 0.04	0.25 ± 0.04	30	N/A	33 ± 3

[a]The mean and error of s_{Ain} from 10 GA runs are given for each case. The dispersity σ_{core} is quantified by the standard deviation in the determined core radius and is presented as a percentage of the averaged core radius size.

3a,c,e, in early generations the $I_{comp}(q)$ of the best individual cannot accurately capture all features in the $I_{exp}(q)$, but with increasing generations, as the GA refines the population during the search for the best candidate in subsequent generations, we see improved agreement in the best individual's $I_{comp}(q)$ with $I_{exp}(q)$ as shown by the corresponding sse value that typically decreases from 1 in early generations to around 0.1−0.01 in later generations (Figure S2). For all three vesicle cases (Figures 3a,c,e), the best individual's $I_{comp}(q)$ converges to the *in silico* $I_{exp}(q)$ within 50 generations of the GA.

The best individual in the GA's final generation holds the values of the vesicle dimensions that produce an $I_{comp}(q)$ closely matched with the input *in silico* $I_{exp}(q)$. As we know the target vesicle dimensions that gave rise to the *in silico* $I_{exp}(q)$, we can compare the GA-determined dimensions to the target vesicle dimensions to validate the GA, and we summarize the relevant target and determined dimensions in Table S1. Since the *in silico* $I_{exp}(q)$ incorporates dispersity in the vesicle R_{core} dimension, we test the capability of CREASE in identifying the correct dimensions of the target vesicle structure when there is a distribution in the size of the vesicles contributing to the input scattering intensity. For each of the three vesicle cases, we perform two sets of GA runs, one where we do not account for dispersity in the vesicle dimensions during the GA's search for the best structure and one where the GA incorporates dispersity into the calculation of the $I_{comp}(q)$ for each individual in the population. In the GA runs without dispersity considerations, we find that the GA-determined dimensions (blue triangles in Figure 3b,d,f) agree well with the individual layer thicknesses and can distinguish the different relative ratios of solvophilic to solvophobic layers for the two target vesicle cases with a thick solvophobic layer (Figure 3d,f). While GA-determined dimensions for the vesicle with the thin B layer case (Figure 3b) cannot resolve the differences in the individual layer thickness as well as the other two cases, all GA-determined dimensions are still within the range of errors coming from multiple runs of GA for each case. This demonstrates the capability of CREASE in correctly identifying the relevant features within a vesicle wall structure. As there is no dispersity in any of the layer thickness dimensions (i.e., t_{Ain}, t_B, and t_{Aout}) during the *in silico* $I_{exp}(q)$ creation, it may not be surprising to see that even when CREASE does not account for any dispersity it still agrees well with the monodisperse target dimensions.

Next, focusing on the vesicle core radius with disperse dimensions, we see that the GA without dispersity still comes close in reaching the target dimension. In the case of the small core radius with thin solvophobic layer (Figure 3b), the GA-determined R_{core} matches the target value, although there is high variation in the GA-determined R_{core} across the 10 GA runs as shown by the large error bar. For the other two cases of vesicle comprised of small core radius with thick solvophobic layer (Figure 3d) and large core radius with thick solvophobic layer (Figure 3f), we see that while the GA runs without considering dispersity give dimensions that come close to the target R_{core} values for the small and large core radii, the GA cannot get an exact match for these systems. The variation in the GA-determined R_{core} across the 10 runs indicates that no one unique solution can describe the disperse vesicles that contribute to the $I_{exp}(q)$. However, the use of multiple GA runs and the associated error is extremely useful information for the user to understand this degeneracy in vesicle dimensions corresponding to a given $I_{exp}(q)$.

In all cases, using the GA with dispersity (orange squares), we can determine the monodisperse target dimensions of layer thicknesses as well as or better than the GA without dispersity runs. In the vesicle with the thin B layer case (Figure 3b), where the GA without dispersity is unable to distinguish the different layer thicknesses, the GA with dispersity can capture the difference in the dimensions for the solvophilic A and solvophobic B layers. As the dispersity in the *in silico* vesicle used to generate the $I_{exp}(q)$ lies in R_{core}, the main improvement of the GA with dispersity over the GA without dispersity is seen in the determined core radius dimensions (Figure 3b). In all three vesicle cases, the average core radius determined over 10 GA runs with dispersity more closely captures the target value and with less variation in the core dimensions across the different GA runs when compared to the GA without dispersity runs. The significantly lower values of sse for GA with dispersity vs GA without dispersity (shown in Figure S2) further confirm the above conclusions. The best structures found by the GA without dispersity have sse converging to higher values, in the range of 0.1−0.9, than that for the GA with dispersity where sse at the end of the GA runs converges to values in the range of 0.01−0.1.

Additionally, the GA with dispersity determines the magnitude of dispersity in the core radius dimension (σ_{core}) used to generate the target vesicle scattering profile during the search for the best structure. For the three cases of vesicles with a 20% dispersity in the target *in silico* vesicle structure (Table 1), the GA-determined dispersity in the core radius gives 24 ± 3%, 18 ± 3%, and 24 ± 2%, which closely follows the variation of 20% in the target core radius dimension.

Finally, we compare our GA-determined results to the corresponding dimensions obtained from traditional analysis involving fits with the core−multishell analytical model of vesicles using the SASVIEW package.[41] The core−multishell analytical model available in SASVIEW[48,49] can capture the individual layer thicknesses in the vesicle wall as shell dimensions (shell 1, 2, and 3), allowing for direct comparison to the GA-determined dimensions for each layer (t_{Ain}, t_B, and t_{Aout}). We note that the analytical model is fit to the *in silico*

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

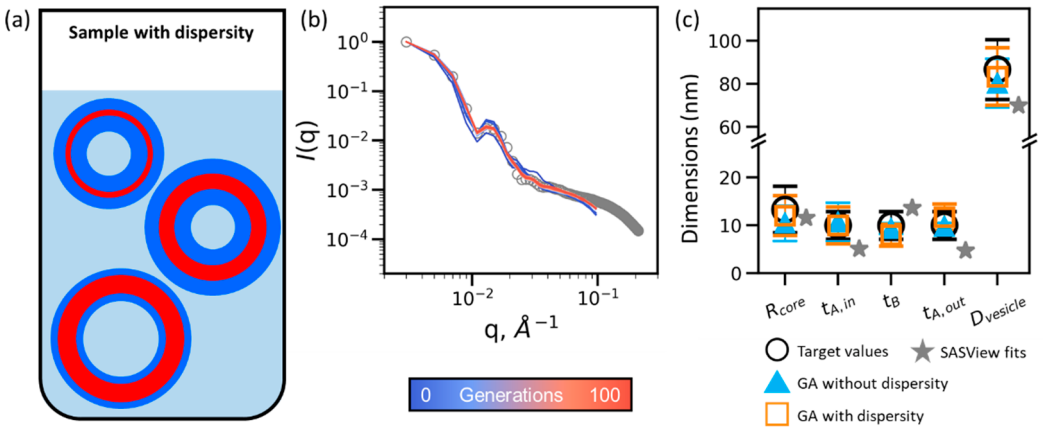


Figure 4. GA analysis on (a) a sample of highly disperse vesicles created using target dimensions that encompass a small core with a thin B layer, a small core with a thick B layer, and a big core with a thick B layer. (b) $I_{comp}(q)$ of the best individual (colored lines) for all generations in one example GA run (with dispersity) compared with the input *in silico* scattering intensity $I_{exp}(q)$ (open circles). Progression from the initial generation to the final generation is shown by increasingly warmer colors. (c) GA-determined dimensions of relevant vesicle features for the cases where dispersity is not accounted for in the GA step (blue triangles) and when dispersity is incorporated into the GA (orange squares). The target dimensions of the *in silico* $I_{exp}(q)$ are shown as open circles. SASVIEW fitting results using the core−multishell model (gray stars) are offset to the right of the target dimensions.

$I_{exp}(q)$ over the same range of q values used in the GA (see SASVIEW fits in Figure S3a−c). Both versions of the GA (with and without dispersity) perform as well as or better than the core−multishell analytical model fit in determining the different vesicle dimensions (Figure 3b,d,f), and in the cases where CREASE outperforms SASVIEW fitting results, the target dimensions all lie well within the standard deviation of the 10 independent GA trials that are performed for each case (Table S1). We also test CREASE on a larger vesicle system (large core with thick A and thin B layers) and like the cases in Figure 3 we observe the enhanced performance of the GA with dispersity over both the GA without considering dispersity and the core−multishell analytical model fitting results (Figure S4).

Figure S2a−d presents the comparison of sse calculated using the log form in eqs 5 for the best SASVIEW fit and best GA fit for all four cases. Comparing the sse values of the best structures determined by SASVIEW and GA, we see that the lowest calculated error does not always correspond to getting the most accurate dimensions. For example, in the case of the vesicle with medium core and thick B layer the GA with dispersity performs the best in obtaining the target dimensions (Figure 3f), but the best structures found by the GA with dispersity all converge to higher sse values of around 0.1 when compared with the calculated sse of 0.01 for the SASVIEW fitting (Figure S2c). The error between the target $I_{exp}(q)$ and the computed or fitted $I(q)$ can guide the search for a better matching candidate within a given method as seen for the GA runs where the sse decreases over generations as the best individual's dimensions approach the target dimensions. However, the sse value is calculated based on fit over an entire q range, and lowering the sse by improving the fit between the target $I_{exp}(q)$ and the computed or fitted $I(q)$ at q values that contribute least to the dimensional information does not improve the accuracy of the predicted dimensions.

It is important to note that regardless of how good the core−multishell analytical model fit is, there is one piece of information that cannot be obtained from the core−multishell analytical model fit that the GA method can determine from the input scattering profile. This is the split in the number of scatterers in the inner vs outer solvophilic layers of the vesicle wall. We note that in the GA step the ratio of number of scatterers in the solvophilic A regions and solvophobic B region is related to the input information about the relative amount of solvophilic to solvophobic chemistry in the polymer solution (see SI Section I.A); however, the relative split of the scatterers in solvophilic A inner and outer layers of the wall is not an input into CREASE and is determined by the GA without prior knowledge of the split fraction. The fraction of scatterers occupying the inner layer of the wall (s_{Ain}) offers additional insight into how solvophilic blocks of the macromolecules are arranged in the assembled structure; this information is currently not available when one fits with the core−multishell analytical model but is very useful to scientists who wish to understand transport of molecular cargo from the core outward through the layers of the vesicle. It is worth noting that a more sophisticated analytical model like the one developed by Brzustowicz and Brunger[50] can determine asymmetric electron densities in the inner and outer solvophilic layers; however, their inherent assumption of Gaussian forms for representing lipid moiety or a region leads to the average molecular density differences between the inner and the outer layers without clear distinction of number and packing of the molecules in those layers. Regardless, it is worth noting that the GA-determined s_{Ain} for the three vesicle cases, incorporating dispersity or not, captures the target s_{Ain} value used to generate the input *in silico* $I_{exp}(q)$ (Table 1).

While the target dimensions in Figure 3 and Figure S4 include dispersity in the core radius for each of the three vesicle cases, the magnitude of the dispersity is limited to 20% of the core radius, and there is no dispersity in the layer thicknesses in the wall. We push the limits of the GA in working with samples with dispersity in all vesicle dimensions, as one may find in realistic experimental samples. We use an averaged $I_{exp}(q)$ created from all three individual disperse $I_{exp}(q)$ from each of the vesicle cases described in Figure 3 (see schematic of this sample with disperse dimensions in Figure 4a). This highly disperse $I_{exp}(q)$ encompasses a large range of values in the core radius and layer thicknesses being sampled, and as such, all target dimensions include uncertainty (R_{core}:

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

13.3 ± 4.8 nm, t_{Ain}: 10.0 ± 2.9 nm, t_B: 10.0 ± 2.9 nm, t_{Aout}: 10.0 ± 2.9 nm, $D_{vesicle}$: 86.7 ± 13.9 nm).

We see that for this highly disperse sample, the GA's best individual's $I_{comp}(q)$ gradually evolves toward the target $I_{exp}(q)$ within 100 generations of the GA (Figure 4b). The best individual's $I_{comp}(q)$ at the end of the GA correctly captures the detailed features in the target $I_{exp}(q)$, such as the small bump at 0.03 $Å^{-1}$, which the fitting using the analytical model fails to pick up (Figure S3e). Most strikingly, for almost all dimensions the GA (with and without dispersity in the core radius) performs significantly better than the analytical model fitting result (Figure 4c), and we see that the SASVIEW results often fall *outside* of the uncertainty in the target dimensions for the individual layer thicknesses while the CREASE for vesicle results all fall *within* the uncertainty in the target dimensions (Table S1). The corresponding sse of the best structure for the GA (with and without dispersity in core radius) and SASVIEW fit are shown in Figure S2e. As seen before, the GA that incorporates dispersity in the core dimension shows an improvement in the determined core radius dimension over the GA without any dispersity consideration. The GA, once again, can capture the split of the solvophilic molecules between the inner and outer solvophilic layers of the vesicle wall and the increased dispersity value of 30% in the core radius dimension (Table 1).

III.B. Molecular Reconstruction of Vesicle Assemblies

Once we have obtained the vesicle dimensions from the first step of CREASE, we use the second step of CREASE to reconstruct the vesicle to determine the molecular level features of concentration profile within the wall and the polymer chain conformations. The reconstruction based on the average dimensions and the split of solvophilic components between the inner and the outer layers of the wall maintains the circular, enclosed wall structure in the cylindrical geometry while retaining the appropriate curvature for the determined size of the vesicle (Figure 5a,d). We simply demonstrate this reconstruction step of CREASE for two representative GA outputs: (i) small core (R_{core} = 10 nm) with thin solvophobic B layer (t_B = 6 nm) and thick solvophilic A layers (t_{Ain} = t_{Aout} = 12 nm) (Figures 5a–c) and (ii) medium core (R_{core} = 20 nm) with thick solvophobic B layer (t_B = 12 nm) and thin solvophilic A layers (t_{Ain} = t_{Aout} = 6 nm) (Figure 5d–f), using two example amphiphilic molecules, peptide–peptide conjugates (Figure 5a) and linear diblock copolymers with a solvophilic-to-solvophobic ratio of A_{18}-*b*-B_{36} (Figure 5d). For this demonstration we use the coarse-grained (CG) model developed for elastin-like-peptide chains conjugated to a collagen-like peptide triple helix (in short ELP-CLP)[43] and a simple bead–spring CG polymer model[42] for the linear diblock copolymer (BCP) chains as described in Molecular Reconstruction step in section II.A.2.

One of the outputs of CREASE's reconstruction step is the monomer concentration profiles as a function of the radial distance from the vesicle center, $C(R)$ (Figure 5b,e), which also confirms that the dimensions output from the GA step are retained after the reconstruction. The GA-determined solvophilic content split between the inner and the outer solvophilic layers in the wall is reflected in the magnitude of the concentration profiles across the two solvophilic and one solvophobic domains. For the ELP-CLP model, the shape of $C(R)$ captures the stiff, rod-like conformation of the solvophilic CLP group in the flattened plateau near the midsections of the

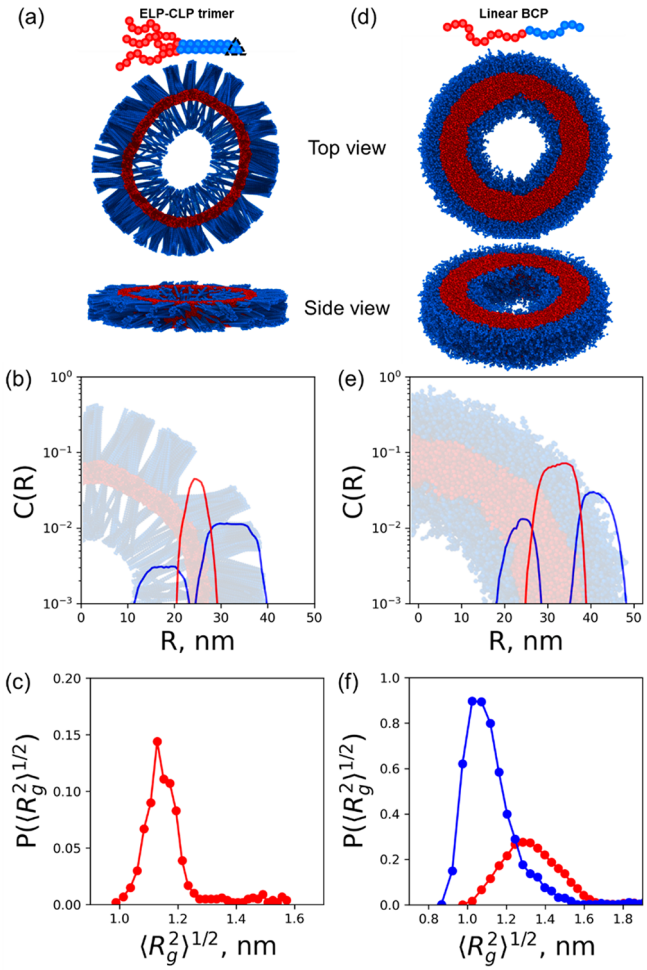


Figure 5. Molecular reconstruction results of a vesicle with a small core with a thin solvophobic layer using the ELP-CLP model (a, b, and c) and a large core with a thick solvophobic layer using a linear diblock copolymer (BCP) model (d, e, and f). (a, d) Visualization of the wall membrane slice from MD simulations. (b, e) Monomer concentration profile of the wall as a function of radial distance from the vesicle center. The solvophilic and solvophobic blocks are shown in blue and red, respectively. (c, f) Distribution of fully flexible chain conformations in the solvophilic and solvophobic layers (blue and red, respectively). Note that for the rigid-rod model of the solvophilic CLP block on the left, the distribution of R_g is meaningless and, thus, not calculated.

CLP layers (Figure 5b), in contrast with the flexible linear BCP model that shows a gradual transition from low to high polymer concentration regions within each solvophilic layer (Figure 5e). Furthermore, one can also see the interfaces of the solvophilic and solvophobic layers through the overlapping regions of the solvophobic (red) and solvophilic (blue) concentration profile curves. The roughness of the interface and the interfacial area between the solvophilic and solvophobic layers can also be quantified using iso-surface analysis[51,52] on the configurations obtained from the reconstructed vesicle.

Another output of CREASE's molecular reconstruction is the distribution of chain conformations, such as the distribution of the radius of gyration, $P[\langle R_g^2\rangle^{1/2}]$ (Figure 5c,f) that is difficult to obtain from analytical model fits which output the average R_g^2. We note that since we use a rigid rod model for the solvophilic CLP block, there is only one fixed R_g

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925–1936

value for the CLP molecules (Figure 5c). In contrast, for the linear diblock copolymers, the reconstructions demonstrate the shift in the average R_g^2 and change in shape of the distribution of the radius of gyration of the solvophilic (blue) and solvophobic (red) blocks (Figure 5f).

With the configurations obtained from the molecular reconstruction, one can characterize additional information beyond the concentration profiles and distribution of chain conformations in each region of the vesicle shown in Figure 5. For example, one can also calculate the orientational order parameter (e.g., S_2) for the semiflexible/stiff polymer blocks[53,54] or liquid crystalline entities in the amphiphilic macromolecules within the different layers of the vesicle. One can expect that the smaller curvature of the outer solvophilic layer as compared to the inner solvophilic layer could impose different confinement effects and bending constraints on such semiflexible/stiff molecules. Further, in the ELP-CLP systems, if we use the more sophisticated coarse-grained model of CLP triple helix that can capture the melting of CLP triple helix into CLP single strands,[55,56] we can also quantify the extent of melted CLP helices in the inner vs outer solvophilic layers. Lastly, the reconstructed equilibrated structure can also provide insight into the membrane bending rigidity and associated thermal fluctuations of the vesicle structure.[57]

The above pieces of information, that are useful in designing vesicles for delivery, are not available to a researcher who is using traditional fitting of the scattering profile with analytical models to interpret structure of vesicles formed in amphiphilic polymer solutions.

IV. CONCLUSION

In this study, we develop and validate the CREASE (computational reverse-engineering analysis for scattering experiments) method for analyzing scattering profiles obtained from macromolecular solutions that exhibit vesicle structures. CREASE is a two-step approach that combines the genetic algorithm (GA) with molecular simulations to capture features in assemblies not easily obtainable through traditional analysis techniques for scattering experiments. We demonstrate how CREASE determines the relevant features and structural dimensions on length scales that range from 100 nm down to a few nanometers for a given scattering profile using a genetic algorithm. The successful molecular reconstruction of vesicles with varying dimensions in the second step of CREASE using molecular simulations and models for two types of macromolecules demonstrates the type of information related to chain-conformation and monomer-level packing that may not be readily available with traditional analysis of scattering results.

Using as input *in silico* $I_{exp}(q)$ created from vesicle structures with known target dimensions, we demonstrate that the GA step of CREASE can determine dimensions that match the target values. By incorporating dispersity during the GA implementation, we can further improve the determined dimensions when the input $I_{exp}(q)$ is created from disperse vesicle dimensions. In all vesicle cases that differ in core radius and layer thicknesses in the vesicle wall that are considered in this paper, CREASE performs as well as or better than the traditional fitting to the $I_{exp}(q)$ using analytical model of vesicles. In the case of disperse samples that have significant variability in core radius and layer thicknesses, CREASE outperforms a traditional analytical model fit in correctly reproducing the target dimensions. Additionally, CREASE quantifies the distribution of the solvophilic block chemistry between the inner and outer layers of the three-layer wall of the vesicle, giving further insight into the morphology of vesicles than possible with traditional analytical fits.

As shown by this study and our previous works,[38−40] the adaptability of CREASE to different assembly shapes makes it a promising technique for studying a wide range of polymer solution assembly structures. The next set of developments for CREASE includes analysis of solvent-processing induced kinetically trapped noncanonical assemblies in polymer solutions[58,59] and temperature-induced anisotropic assemblies of biopolymers,[60] which have no existing analytical models. Furthermore, extension of CREASE to concentrated solutions of such assemblies will also include the calculation of $S_{MM}(q)$ (where M is a "micelle") to account for interaction between assembled micelles; in such cases CREASE will include two optimization loops, one for form factor and one for structure factor. We direct readers who are interested in using CREASE for their scattering analysis to visit the page [https://github.com/arthijayaraman-lab/crease_ga], which holds the open-source CREASE-GA code and a step-by-step tutorial for CREASE-GA applied to vesicles.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.1c00305.

Additional details of molecular simulations and models used in the reconstruction step, analytical fits using SASVIEW, and further tests that we have performed (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Arthi Jayaraman − *Colburn Laboratory, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States; Department of Materials Science and Engineering, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0002-5295-4581; Email: arthij@udel.edu

Authors

Ziyu Ye − *Colburn Laboratory, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States*

Zijie Wu − *Colburn Laboratory, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.1c00305

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the National Science Foundation, Grant Number NSF ENG 2023668, for the financial support. The authors are grateful to Mikhail Zhernenkov at Brookhaven National Laboratory for his expert advice and guidance on the usage of SASVIEW. The simulations in this paper are

performed on the high-performance computing resources of the Farber supercomputing cluster maintained by the Information Technologies (IT) department at the University of Delaware.

REFERENCES

(1) Jiao, D.; Geng, J.; Loh, X. J.; Das, D.; Lee, T.-C.; Scherman, O. A. Supramolecular Peptide Amphiphile Vesicles through Host−Guest Complexation. *Angew. Chem., Int. Ed.* **2012**, *51* (38), 9633−9637.
(2) Gaitzsch, J.; Huang, X.; Voit, B. Engineering Functional Polymer Capsules toward Smart Nanoreactors. *Chem. Rev.* **2016**, *116* (3), 1053−1093.
(3) Al-Ahmady, Z.; Kostarelos, K. Chemical Components for the Design of Temperature-Responsive Vesicles as Cancer Therapeutics. *Chem. Rev.* **2016**, *116* (6), 3883−3918.
(4) Pick, H.; Alves, A. C.; Vogel, H. Single-Vesicle Assays Using Liposomes and Cell-Derived Vesicles: From Modeling Complex Membrane Processes to Synthetic Biology and Biomedical Applications. *Chem. Rev.* **2018**, *118* (18), 8598−8654.
(5) Nishimura, T.; Hirose, S.; Sasaki, Y.; Akiyoshi, K. Substrate-Sorting Nanoreactors Based on Permeable Peptide Polymer Vesicles and Hybrid Liposomes with Synthetic Macromolecular Channels. *J. Am. Chem. Soc.* **2020**, *142* (1), 154−161.
(6) Gorgoll, R. M.; Harano, K.; Nakamura, E. Nanoscale Control of Polymer Assembly on a Synthetic Catalyst−Bilayer System. *J. Am. Chem. Soc.* **2016**, *138* (30), 9675−9681.
(7) Discher, D. E.; Eisenberg, A. Polymer Vesicles. *Science* **2002**, *297* (5583), 967.
(8) Koide, A.; Kishimura, A.; Osada, K.; Jang, W.-D.; Yamasaki, Y.; Kataoka, K. Semipermeable Polymer Vesicle (PICsome) Self-Assembled in Aqueous Medium from a Pair of Oppositely Charged Block Copolymers: Physiologically Stable Micro-/Nanocontainers of Water-Soluble Macromolecules. *J. Am. Chem. Soc.* **2006**, *128* (18), 5988−5989.
(9) Luo, T.; Kiick, K. L. Noncovalent Modulation of the Inverse Temperature Transition and Self-Assembly of Elastin-b-Collagen-like Peptide Bioconjugates. *J. Am. Chem. Soc.* **2015**, *137* (49), 15362−15365.
(10) Rahman, M. M.; Ueda, M.; Hirose, T.; Ito, Y. Spontaneous Formation of Gating Lipid Domain in Uniform-Size Peptide Vesicles for Controlled Release. *J. Am. Chem. Soc.* **2018**, *140* (51), 17956−17961.
(11) Qin, J.; Sloppy, J. D.; Kiick, K. L. Fine structural tuning of the assembly of ECM peptide conjugates via slight sequence modifications. *Science Advances* **2020**, *6* (41), eabd3033.
(12) McCoy, T. M.; Marlow, J. B.; Armstrong, A. J.; Clulow, A. J.; Garvey, C. J.; Manohar, M.; Darwish, T. A.; Boyd, B. J.; Routh, A. F.; Tabor, R. F. Spontaneous Self-Assembly of Thermoresponsive Vesicles Using a Zwitterionic and an Anionic Surfactant. *Biomacromolecules* **2020**, *21* (11), 4569−4576.
(13) Nishimura, T.; Shishi, S.; Sasaki, Y.; Akiyoshi, K. Thermoresponsive Polysaccharide Graft Polymer Vesicles with Tunable Size and Structural Memory. *J. Am. Chem. Soc.* **2020**, *142* (27), 11784−11790.
(14) Georgiou, P. G.; Marton, H. L.; Baker, A. N.; Congdon, T. R.; Whale, T. F.; Gibson, M. I. Polymer Self-Assembly Induced Enhancement of Ice Recrystallization Inhibition. *J. Am. Chem. Soc.* **2021**, *143* (19), 7449−7461.
(15) Franken, L. E.; Boekema, E. J.; Stuart, M. C. A. Transmission Electron Microscopy as a Tool for the Characterization of Soft Materials: Application and Interpretation. *Advanced Science* **2017**, *4* (5), 1600476.
(16) Friedrich, H.; Frederik, P. M.; de With, G.; Sommerdijk, N. A. J. M. Imaging of Self-Assembled Structures: Interpretation of TEM and Cryo-TEM Images. *Angew. Chem., Int. Ed.* **2010**, *49* (43), 7850−7858.
(17) Patterson, J. P.; Kelley, E. G.; Murphy, R. P.; Moughton, A. O.; Robin, M. P.; Lu, A.; Colombani, O.; Chassenieux, C.; Cheung, D.; Sullivan, M. O.; Epps, T. H.; O'Reilly, R. K. Structural Characterization of Amphiphilic Homopolymer Micelles Using Light Scattering, SANS, and Cryo-TEM. *Macromolecules* **2013**, *46* (15), 6319−6325.
(18) Li, Z.; Van Zee, N. J.; Bates, F. S.; Lodge, T. P. Polymer Nanogels as Reservoirs To Inhibit Hydrophobic Drug Crystallization. *ACS Nano* **2019**, *13* (2), 1232−1243.
(19) Ghezzi, M.; Pescina, S.; Padula, C.; Santi, P.; Del Favero, E.; Cantù, L.; Nicoli, S. Polymeric micelles in drug delivery: An insight of the techniques for their characterization and assessment in biorelevant conditions. *J. Controlled Release* **2021**, *332*, 312−336.
(20) Pedersen, J. S.; Hamley, I. W.; Ryu, C. Y.; Lodge, T. P. Contrast Variation Small-Angle Neutron Scattering Study of the Structure of Block Copolymer Micelles in a Slightly Selective Solvent at Semidilute Concentrations. *Macromolecules* **2000**, *33* (2), 542−550.
(21) Pedersen, J. S.; Svaneborg, C.; Almdal, K.; Hamley, I. W.; Young, R. N. A Small-Angle Neutron and X-ray Contrast Variation Scattering Study of the Structure of Block Copolymer Micelles: Corona Shape and Excluded Volume Interactions. *Macromolecules* **2003**, *36* (2), 416−433.
(22) Hammouda, B. SANS from Polymers—Review of the Recent Literature. *Polym. Rev.* **2010**, *50* (1), 14−39.
(23) Gilroy, J. B.; Rupar, P. A.; Whittell, G. R.; Chabanne, L.; Terrill, N. J.; Winnik, M. A.; Manners, I.; Richardson, R. M. Probing the Structure of the Crystalline Core of Field-Aligned, Monodisperse, Cylindrical Polyisoprene-block-Polyferrocenylsilane Micelles in Solution Using Synchrotron Small- and Wide-Angle X-ray Scattering. *J. Am. Chem. Soc.* **2011**, *133* (42), 17056−17062.
(24) Mable, C. J.; Gibson, R. R.; Prevost, S.; McKenzie, B. E.; Mykhaylyk, O. O.; Armes, S. P. Loading of Silica Nanoparticles in Block Copolymer Vesicles during Polymerization-Induced Self-Assembly: Encapsulation Efficiency and Thermally Triggered Release. *J. Am. Chem. Soc.* **2015**, *137* (51), 16098−16108.
(25) Wang, E.; Lu, J.; Bates, F. S.; Lodge, T. P. Effect of Corona Block Length on the Structure and Chain Exchange Kinetics of Block Copolymer Micelles. *Macromolecules* **2018**, *51* (10), 3563−3571.
(26) Pokorski, J. K.; Hore, M. J. A. Structural characterization of protein−polymer conjugates for biomedical applications with small-angle scattering. *Curr. Opin. Colloid Interface Sci.* **2019**, *42*, 157−168.
(27) Zhao, D.; Ma, Y.; Wang, E.; Lodge, T. P. Micellization of Binary Diblock Co-polymer Mixtures in an Ionic Liquid. *Macromolecules* **2019**, *52* (12), 4729−4738.
(28) Zhao, D.; Wang, E.; Lodge, T. P. Hybridization of a Bimodal Distribution of Copolymer Micelles. *Macromolecules* **2020**, *53* (18), 7705−7716.
(29) Kim, S.; Cho, Y.; Kim, J. H.; Song, S.; Lim, J.; Choi, S.-H.; Char, K. Structural Analysis of Bottlebrush Block Copolymer Micelles Using Small-Angle X-ray Scattering. *ACS Macro Lett.* **2020**, *9* (9), 1261−1266.
(30) Sochor, B.; Düdükcü, Ö.; Lübtow, M. M.; Schummer, B.; Jaksch, S.; Luxenhofer, R. Probing the Complex Loading-Dependent Structural Changes in Ultrahigh Drug-Loaded Polymer Micelles by Small-Angle Neutron Scattering. *Langmuir* **2020**, *36* (13), 3494−3503.
(31) Qian, S.; Sharma, V. K.; Clifton, L. A. Understanding the Structure and Dynamics of Complex Biomembrane Interactions by Neutron Scattering Techniques. *Langmuir* **2020**, *36* (50), 15189−15211.
(32) Wei, Y.; Hore, M. J. A. Characterizing polymer structure with small-angle neutron scattering: A Tutorial. *J. Appl. Phys.* **2021**, *129* (17), 171101.
(33) Bang, J.; Jain, S.; Li, Z.; Lodge, T. P.; Pedersen, J. S.; Kesselman, E.; Talmon, Y. Sphere, Cylinder, and Vesicle Nanoaggregates in Poly(styrene-b-isoprene) Diblock Copolymer Solutions. *Macromolecules* **2006**, *39* (3), 1199−1208.
(34) Bergström, M.; Pedersen, J. S.; Schurtenberger, P.; Egelhaaf, S. U. Small-Angle Neutron Scattering (SANS) Study of Vesicles and Lamellar Sheets Formed from Mixtures of an Anionic and a Cationic Surfactant. *J. Phys. Chem. B* **1999**, *103* (45), 9888−9897.
(35) Vogtt, K.; Jeworrek, C.; Garamus, V. M.; Winter, R. Microdomains in Lipid Vesicles: Structure and Distribution Assessed

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936

by Small-Angle Neutron Scattering. *J. Phys. Chem. B* **2010**, *114* (16), 5643−5648.
(36) Kučerka, N.; Nagle, J. F.; Feller, S. E.; Balgavý, P. Models to analyze small-angle neutron scattering from unilamellar lipid vesicles. *Phys. Rev. E* **2004**, *69* (5), No. 051903.
(37) Eicher, B.; Heberle, F. A.; Marquardt, D.; Rechberger, G. N.; Katsaras, J.; Pabst, G. Joint small-angle X-ray and neutron scattering data analysis of asymmetric lipid vesicles. *J. Appl. Crystallogr.* **2017**, *50* (2), 419−429.
(38) Beltran-Villegas, D. J.; Wessels, M. G.; Lee, J. Y.; Song, Y.; Wooley, K. L.; Pochan, D. J.; Jayaraman, A. Computational Reverse-Engineering Analysis for Scattering Experiments on Amphiphilic Block Polymer Solutions. *J. Am. Chem. Soc.* **2019**, *141* (37), 14916−14930.
(39) Lee, J. Y.; Song, Y.; Wessels, M. G.; Jayaraman, A.; Wooley, K. L.; Pochan, D. J. Hierarchical Self-Assembly of Poly(d-glucose carbonate) Amphiphilic Block Copolymers in Mixed Solvents. *Macromolecules* **2020**, *53* (19), 8581−8591.
(40) Wessels, M. G.; Jayaraman, A. Computational Reverse-Engineering Analysis of Scattering Experiments (CREASE) on Amphiphilic Block Polymer Solutions: Cylindrical and Fibrillar Assembly. *Macromolecules* **2021**, *54* (2), 783−796.
(41) http://www.sasview.org (accessed June 3, 2021).
(42) Grest, G. S.; Kremer, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33* (5), 3628−3631.
(43) Condon, J. E.; Martin, T. B.; Jayaraman, A. Effect of conjugation on phase transitions in thermoresponsive polymers: an atomistic and coarse-grained simulation study. *Soft Matter* **2017**, *13* (16), 2907−2918.
(44) Jones, J. E. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond. A* **1924**, *106* (738), 441−462.
(45) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1−19.
(46) Kamberaj, H.; Low, R. J.; Neal, M. P. Time reversible and symplectic integrators for molecular dynamics simulations of rigid molecules. *J. Chem. Phys.* **2005**, *122* (22), 224114.
(47) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54* (12), 5237−5247.
(48) Guinier, A.; Fournet, G.; Walker, C. B.; Yudowitch, K. L. *Small Angle Scattering of X-rays*; John Wiley and Sons: New York, 1955.
(49) Feigin, L. A.; Svergun, D. I. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*; Plenum Press: New York, 1987.
(50) Brzustowicz, M. R.; Brunger, A. T. X-ray scattering from unilamellar lipid vesicles. *J. Appl. Crystallogr.* **2005**, *38* (1), 126−131.
(51) Lorensen, W. E.; Cline, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.* **1987**, *21* (4), 163−169.
(52) Prasad, I.; Jinnai, H.; Ho, R.-M.; Thomas, E. L.; Grason, G. M. Anatomy of triply-periodic network assemblies: characterizing skeletal and inter-domain surface geometry of block copolymer gyroids. *Soft Matter* **2018**, *14* (18), 3612−3623.
(53) Milchev, A.; Egorov, S. A.; Nikoubashman, A.; Binder, K. Adsorption and structure formation of semiflexible polymers on spherical surfaces. *Polymer* **2018**, *145*, 463−472.
(54) Khadilkar, M. R.; Nikoubashman, A. Self-assembly of semiflexible polymers confined to thin spherical shells. *Soft Matter* **2018**, *14* (33), 6903−6911.
(55) Condon, J. E.; Jayaraman, A. Development of a Coarse-Grained Model of Collagen-Like Peptide (CLP) for Studies of CLP Triple Helix Melting. *J. Phys. Chem. B* **2018**, *122* (6), 1929−1939.
(56) Hilderbrand, A. M.; Taylor, P. A.; Stanzione, F.; LaRue, M.; Guo, C.; Jayaraman, A.; Kloxin, A. M. Combining simulations and experiments for the molecular engineering of multifunctional collagen mimetic peptide-based materials. *Soft Matter* **2021**, *17* (7), 1985−1998.
(57) Chappa, V.; Smirnova, Y.; Komorowski, K.; Muller, M.; Salditt, T. The effect of polydispersity, shape fluctuations and curvature on small unilamellar vesicle small-angle X-ray scattering curves. *J. Appl. Crystallogr.* **2021**, *54* (2), 557−568.
(58) Kim, K. T.; Zhu, J.; Meeuwissen, S. A.; Cornelissen, J. J. L. M.; Pochan, D. J.; Nolte, R. J. M.; van Hest, J. C. M. Polymersome Stomatocytes: Controlled Shape Transformation in Polymer Vesicles. *J. Am. Chem. Soc.* **2010**, *132* (36), 12522−12524.
(59) Vena, M. P.; de Moor, D.; Ianiro, A.; Tuinier, R.; Patterson, J. P. Kinetic state diagrams for a highly asymmetric block copolymer assembled in solution. *Soft Matter* **2021**, *17* (4), 1084−1090.
(60) Cui, Y.; Lawoko, M.; Svagan, A. J. High Value Use of Technical Lignin. Fractionated Lignin Enables Facile Synthesis of Microcapsules with Various Shapes: Hemisphere, Bowl, Mini-tablets, or Spheres with Single Holes. *ACS Sustainable Chem. Eng.* **2020**, *8* (35), 13282−13291.

https://doi.org/10.1021/jacsau.1c00305
JACS Au 2021, 1, 1925−1936