

Computational Reverse-Engineering Analysis for Scattering Experiments on Amphiphilic Block Polymer Solutions

Daniel J. Beltran-Villegas,[†] Michiel G. Wessels,[†] Jee Young Lee,[‡] Yue Song,[§] Karen L. Wooley,^{*,§} Darrin J. Pochan,^{*,‡} and Arthi Jayaraman^{*,‡,§}

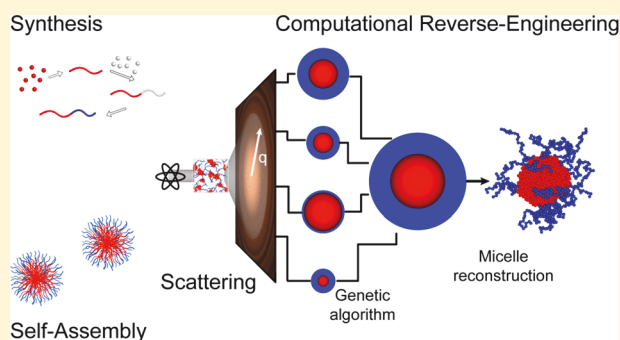
[†]Department of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy Street, Colburn Laboratory, Newark, Delaware 19716, United States

[‡]Department of Materials Science and Engineering, University of Delaware, 201 DuPont Hall, Newark, Delaware 19716, United States

[§]Departments of Chemistry, Chemical Engineering, and Materials Science & Engineering, Texas A&M University, College Station, Texas 77842, United States

Supporting Information

ABSTRACT: In this paper, we present a computational reverse-engineering analysis for scattering experiments (CREASE) based on genetic algorithms and molecular simulation to analyze the structure within self-assembled amphiphilic polymer solutions. For a given input comprised of scattering intensity profiles and information about the amphiphilic polymers in solution, CREASE outputs the structure of the self-assembled micelles (e.g., core and corona diameters, aggregation number) as well as the conformations of the amphiphilic polymer chains in the micelle (e.g., blocks' radii of gyration, chain radii of gyration, monomer concentration profiles). First, we demonstrate CREASE's ability to reverse-engineer self-assembled nanostructures for scattering profiles obtained from molecular simulations (or *in silico* experiments) of generic coarse-grained bead-spring polymer chains in an implicit solvent. We then present CREASE's outputs for scattering profiles obtained from small-angle neutron scattering (SANS) experiments of poly(*D*-glucose carbonate) block copolymers in solution that exhibit assembly into spherical nanoparticles. The success of this method is demonstrated by its ability to replicate, quantitatively, the results from *in silico* experiments and by the agreement in micelle core and corona sizes obtained from microscopy of the *in vitro* solutions. The primary strength of CREASE is its ability to analyze scattering profiles without an off-the-shelf scattering model and the ability to provide chain and monomer level structural information that is otherwise difficult to obtain from scattering and microscopy alone.



I. INTRODUCTION

Over the past two decades, nanoscience has evolved from a vision to a nanotechnological reality, yet only recently has the capability of theory and computation evolved to the point of being able to understand and predict composition–structure–property phenomena to guide experimental designs. One particular field where this powerful synergy of theory, simulation, and experiments has been harnessed is in self-assembly of amphiphilic block copolymers (BCPs) in solution.^{1–7} BCPs self-assemble in solution to form well-defined nanostructures that have been used in a broad range of applications within healthcare and medicine,⁸ sensing,⁹ separations,¹⁰ electronics,¹¹ and environmental sciences.^{12,13} To tailor the assembled structures for these applications, many computational and experimental studies have focused on understanding how the polymer design (e.g., polymer chemistry,^{14,15} composition,^{16–19} and architecture^{20,21}) and solution conditions (e.g., temperature, solvent,^{14,22,23} pH,^{24–26}

and small-molecule addition²⁷) impact the assembly. A critical step to accomplish this goal of designing BCPs and optimizing nanostructures involves the unambiguous characterization of the self-assembled structure. The next frontier involves harnessing the power of simulations to analyze in depth the molecular details that are contained within the data from traditional experimental characterization techniques.

To characterize the structure of BCP micelles, experimentalists rely on microscopy and scattering techniques, such as small-angle X-ray and neutron scattering. In particular, small-angle neutron scattering (SANS) is a powerful tool for probing BCP-assembled structures at various length scales from local chain conformations (e.g., radius of gyration) to micelle dimensions (e.g., aggregation number, micelle size). The measured quantity from a SANS experiment is the scattering

Received: July 26, 2019

Published: September 9, 2019



Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE)

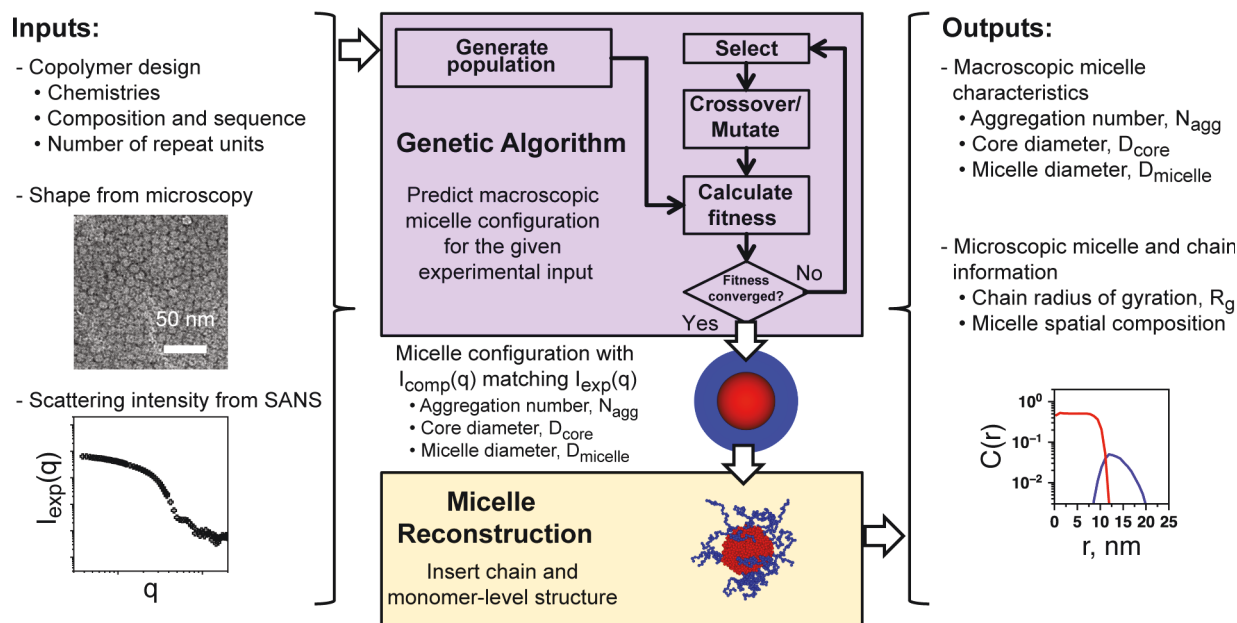


Figure 1. Schematic flow diagram of computational reverse-engineering analysis for scattering experiments.

intensity, $I(q)$, of neutrons elastically colliding with the sample as a function of the wavevector, q , that encodes information on correlations between different scattering species in reciprocal space. These SANS results complement the real-space but short-length and low-contrast information provided by transmission electron microscopy (TEM). SANS is particularly useful in the analysis of BCP micelles because different chemical blocks can be selectively deuterated to contrast-match with the solvent and highlight chain- and micelle-level features of the noncontrast-matched components,²⁸ allowing for selective core or corona signals.²⁹

Interpretation of $I(q)$ typically relies on (mathematical) models that have been developed for some of the canonical shapes of the assembled micelles.²⁹ In particular, models for BCP micelles developed by Pedersen and co-workers³⁰ rely on assumptions regarding the nature and shape²⁹ of the core (e.g., spherical shape, with possible swelling by solvents²²), corona chain conformations (e.g., Gaussian^{30,31} and semiflexible with excluded volume effects³¹), and micelle size polydispersity effects.^{30,32,33} Correct usage and fitting of $I(q)$ with these models requires extensive experience, and correct interpretation of the fit results may not be obvious to a nonexpert. Additionally, chain-level structural features from these fits are contingent on the assumptions (e.g., Gaussian chain conformations) that the model is based on. To reliably interpret SANS measurements, one needs an optimization method that is able to tie the features of $I(q)$ directly to the micelle, chain, and monomer features without needing off-the-shelf analytical models. To address this need, we present in this article an approach based on genetic algorithms and coarse-grained molecular simulations.

A genetic algorithm (GA) is a robust global optimization method^{34,35} that has been used in materials discovery^{36,37} and for designing polymeric,^{38–42} colloidal,⁴³ and inorganic^{44–47} materials. GA is inspired by evolutionary processes where the

“fitness”—a function that encodes the desirability of a set of parameters toward a goal—of a population of individuals is maximized by selecting “good” individuals and applying genetic operators on selected individuals to improve the overall population fitness while maintaining diversity in order to avoid local optima.^{34,35} This last feature makes it useful in problems with inherently complex fitness functions, such as the problem of extracting micelle properties from $I(q)$.

Coarse-grained molecular dynamics (CGMD) simulations have been used extensively to understand the fundamentals of BCP assembly in solution, shown in review articles^{48,49} as well as our recent studies,^{50–55} including the relationship between chain composition, sequence (e.g., diblock, triblock), architecture, concentration, and solvent quality on micelle shape, size, and chain conformations. With an appropriate level of coarse graining, CGMD simulations capture relevant length and time scales of the macromolecular solutions to shed light on both the (relatively) macroscopic-assembled structure and microscopic chain-level features that are usually hard to observe directly from experiments, including chain conformations and spatial arrangement of the different monomers in a micelle.

We introduce in this paper a two-stage computational method based first on a GA stage for evaluation of micelle-level structure from the $I(q)$ and then a second stage involving CGMD simulation to add finer resolution to micelle-level structure and attain chain- and monomer-level structural detail. We name this method computational reverse-engineering analysis for scattering experiments (CREASE). In this paper, we present the details and implementation of CREASE for the specific case of spherical BCP micelles. First, we evaluate the method against simulated experiments, where both micelle-level and chain-level features are known, to test the predictive power and understand possible drawbacks of CREASE. We then proceed to evaluate CREASE with experimental SANS

measurements of poly(D-glucose carbonate) (PGC)-based amphiphilic diblock copolymer micelles. Beyond this proof-of-concept, in future work, we expect CREASE to be extended to other systems with polymer chain assembly by using different global optimization methods or become integrated with different machine learning algorithms to improve the level of structural detail and the efficiency of the parameter space explored. Additionally, different levels of coarse-grained models or atomistic simulations could be used to extract the desired level of structural detail.

The breadth of the work presented in this paper extends from the synthesis of natural product-derived and recyclable glucose carbonates to their supramolecular assembly in solution, as analyzed experimentally, with a focus on method development that allows for reverse engineering of simulated nanoscopic assemblies. The future promise of this dual experimental–computational effort is in the advanced understanding of the global and local molecular details within complex chemical systems.

II. METHODS

II.A. CREASE Method Description. In this section, we present the key steps in CREASE. Inputs to CREASE include the scattering experiment measurement, information about the BCP chemistry used, and the shape of the BCP assembly. This input information is fed into a *genetic algorithm* (GA) step that determines macroscopic features of the assembly (e.g., micelle dimensions, number of BCP chains in a micelle). This output from the GA step is then fed to a *micelle reconstruction* step where coarse-grained molecular dynamics (CGMD) simulations are used to provide microscopic features of the assembly (e.g., BCP chain conformations, spatial distribution of the monomers in the micelle). Put together, for a given scattering intensity profile and BCP design, CREASE provides structural information starting from the length scale of a monomer up to the entire micelle. Figure 1 shows this overall flow of information and these key steps in CREASE.

II.A.1. Input Information. The input parameters for CREASE include the BCP design, the experimentally obtained scattering intensity, $I_{\text{exp}}(q)$, where q is the wavevector, and the shape of the micelle that can be obtained from an independent measurement, such as microscopy. The BCP design includes: (1) the chemistry of the solvophilic A and solvophobic B monomers, the molar masses of A and B repeat units, M_A and M_B , and their contour lengths, $l_{\text{mono,A}}$ and $l_{\text{mono,B}}$; (2) the BCP composition, f_A , i.e., the fraction of repeat units that correspond to the A block chemistry, and the sequence in which the A and B chemistries are arranged, e.g., AB or ABA; (3) the total molecular weight or number of repeat units, N , and (4) the density of the polymer in the solvophobic B block, ρ_p , taken as the bulk density of the B polymer chemistry at the same conditions, temperature, and pressure as the micelle.

In the present work, we focus on *spherical* micelles comprised of a spherical micelle core of a B block surrounded by a shell (corona) of an A block, as a proof of concept of the method. The method can be extended to other micelle shapes (e.g., cylinders, vesicles) as well. As stated above, the shape of the micelle is an input to CREASE. One could estimate the dimensionality and/or shape of the micelle based on the Porod exponent and exponents of the intermediate q -range.⁵⁶ However, there are multiple ways to interpret the Porod or intermediate q -range exponents as the exponents are not necessarily unique to each shape; for example, both rectangular slabs and cylindrical structures have a possible limiting Porod exponent of 4 at high q , indicating a sharp interface between their respective three-dimensional shape and the solvent. Additionally, the fitted exponent may not be a whole number, due to experimental uncertainty or sample inhomogeneity. Therefore, to minimize the uncertainty involved with scattering data, we recommend an independent measurement (e.g., microscopy) for estimating micelle shape.

II.A.2. Genetic Algorithm (GA). We use a GA to find the macroscopic micelle properties that result in a computed scattering intensity, $I_{\text{comp}}(q)$, that best matches $I_{\text{exp}}(q)$. The algorithm starts with a random population of potential macroscopic micelle characteristics (individuals). “Fitness” is evaluated for all the individuals of the population to determine which individuals result in $I_{\text{comp}}(q)$ closer to $I_{\text{exp}}(q)$; i.e., an individual with a high fitness has a $I_{\text{comp}}(q)$ that is a close match to $I_{\text{exp}}(q)$. In the selection step we pick the best individuals from those available based on their fitness; a solution with high fitness has a higher probability of being selected. Some of the selected individuals undergo crossover or mutation, and the new population (a new generation) advances to the next iteration of the algorithm. The procedure is repeated for several generations until the average population fitness converges to a plateau value. The final individual reported corresponds to the best match between $I_{\text{comp}}(q)$ and $I_{\text{exp}}(q)$ (highest fitness) from the last generation. In the following passages we go through the various steps in the GA in more detail.

Step 1: Generating Individuals of a Population and Calculating Computed Scattering Intensity, $I_{\text{comp}}(q)$. The set of macroscopic micelle characteristics that impact $I_{\text{comp}}(q)$ includes the input parameters to CREASE (section II.A.1.)—A and B polymer chemistries, f_A , N , and ρ_p —and structural parameters that are calculated during the GA—the number of chains per micelle or the micelle aggregation number (N_{agg}), the diameter of the micelle core (D_{core}), the extent of the corona (E_{corona}), the total micelle diameter (D_{micelle}), and the background scattering intensity ($I_{\text{background}}$). E_{corona} is a parameter between 0 (completely collapsed corona) and 1 (completely extended corona). As shown in the equations below, N_{agg} and E_{corona} are the two independent parameters that can uniquely define a spherical micelle. The micelle core diameter D_{core} is

$$D_{\text{core}} = \left(\frac{6 N_{\text{agg}} N (1 - f_A) M_B}{\pi N_A \rho_p} \right)^{1/3} \quad (1)$$

where N_A is the Avogadro number. The total micelle diameter is

$$D_{\text{micelle}} = D_{\text{core}} + 2E_{\text{corona}}(N f_A l_{\text{mono,A}}) \quad (2)$$

The maximum distance the corona can extend past the micelle core is the contour length of the solvophilic block of one chain.

At the start of a GA run we generate a population of 100 random individuals given by the set N_{agg} in the range [2, 100], E_{corona} in the range [0, 1], and $I_{\text{background}}$ in the range [10^{-5} , 1]. When choosing the population size, it is important to have a population large enough to allow for significant genetic diversity while balancing the increased computational power required to run a larger population. While the choice of population size is problem-dependent, in our case the consistency between duplicate runs indicates that we have an adequate population size for our systems. For each individual of the population, we need to calculate the $I_{\text{comp}}(q)$ based on the set of initial parameters. To do that, we first calculate the intramolecular structure factor of a micelle, $\omega(q)$, as

$$\omega(q) = \left\langle \frac{1}{N_A + N_B} \sum_{i=1}^{N_A+N_B} \sum_{j=1}^{N_A+N_B} \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle \quad (3)$$

where r_{ij} is the distance between point scatterers i and j , and N_A and N_B are the number of point scatterers of blocks A and B, respectively. We note that eq 3 includes the scattering of both solvophilic and solvophobic blocks and is valid for experimental cases where both chemistries scatter neutrons similarly with respect to the solvent and background. In the case where one species is contrast-matched with the solvent, that species would not count toward $\omega(q)$, and eq 3 would need to be changed accordingly. To calculate $\omega(q)$ with eq 3, we place N_B point scatterers randomly in the micelle core and N_A point scatterers in the micelle corona for a given N_{agg} , D_{core} , and D_{micelle} . For each individual's N_{agg} , D_{core} , and D_{micelle} , we generate three independent, randomly generated configurations of the scatterers to get an average $\omega(q)$. We note that the calculation of $\omega(q)$ in eq 3 is computationally expensive and is the rate-limiting step of the GA. To

increase the speed of this calculation, we do not have one point scatterer per atom but instead treat BCP as an 8-point scatterer sequence of beads, each of size $\sigma_{GA,A}$ or $\sigma_{GA,B}$ (for the A and B block, respectively) given by

$$\sigma_{GA,A} = \frac{Nl_{\text{mono},A}}{8} \quad (4)$$

and

$$\sigma_{GA,B} = \frac{Nl_{\text{mono},B}}{8} \quad (5)$$

The number of point scatterers in a micelle is then $N_A = 8f_A N_{\text{agg}}$ and $N_B = 8(1 - f_A) N_{\text{agg}}$. The choice of 8 scatterer beads is arbitrary and can indeed be any number that the user wishes to use while balancing the level of detail used to represent the polymer and the computational expense for this calculation of $\omega(q)$. We tested a different number of scatterers per chain, namely, 4, 8, and 12, and found that 4 scatterers per chain result in a $\sigma_{GA,B}$ similar to D_{core} , which reduces the ability to correctly resolve D_{core} and results in a poor GA performance. For 8 and 12 scatterers per chain our results are not significantly impacted by the number of scatterers for these cases (see Section S.1). Besides choosing scatterer size and number of scatterers, there are other ways to speed up/avoid this rate-limiting calculation of $\omega(q)$ in eq 3. These include making use of tabulated $\omega(q)$ for specific input parameters, treating the calculation semi-analytically when certain structural features (e.g., shape) are well-known and have corresponding equations for $\omega(q)$, and incorporating machine learning algorithms to bypass or speed up the fitness calculation after enough generations have passed to train the algorithm.^{57,58}

After calculating the $\omega(q)$, the micelle form factor is calculated as

$$F_M^2(q) = \sum_{\alpha \in \{A,B\}} \sum_{\beta \in \{A,B\}} b_\alpha b_\beta F_{GA,\alpha}(q) F_{GA,\beta}(q) \omega(q) \quad (6)$$

where b_A and b_B are the scattering lengths of A and B scatterers, respectively, and $F_{GA,A}(q)$ and $F_{GA,B}(q)$ are the spherical form factors of spheres of diameter $\sigma_{GA,A}$ and $\sigma_{GA,B}$, respectively, given by

$$F_{GA,\alpha}(q) = \frac{3[\sin(q0.5\sigma_{GA,\alpha}) - q0.5\sigma_{GA,\alpha} \cos(q0.5\sigma_{GA,\alpha})]}{(q0.5\sigma_{GA,\alpha})^3} \quad (7)$$

where α stands for A or B. Finally, the computed scattering intensity $I_{\text{comp}}(q)$ is given by

$$I_{\text{comp}}(q) = F_M^2(q) S_{\text{MM}}(q) + I_{\text{background}} \quad (8)$$

where $S_{\text{MM}}(q)$ is the micelle–micelle structure factor and $I_{\text{background}}$ is the experimental background scattering intensity. In the present study, we focus on dilute polymer solutions that result in negligible micelle–micelle interactions and thus, $S_{\text{MM}}(q) = 1$. For concentrated solutions with micelle–micelle interactions, $S_{\text{MM}}(q)$ depends on the way micelles interact with each other. A simple assumption is to treat micelles as hard spheres of effective diameter $D_{\text{micelle,eff}}$. $D_{\text{micelle,eff}} = D_{\text{micelle}}$ for neutral BCPs in good solvent conditions and $D_{\text{micelle,eff}} > D_{\text{micelle}}$ for charged BCPs with an electrostatics-derived repulsion between micelles. Using hard spheres of $D_{\text{micelle,eff}}$ at the appropriate micelle concentration, one can calculate $S_{\text{MM}}(q)$ either from liquid-state theory^{59,60} or from a radial distribution function obtained from configurations sampled in molecular simulations.⁶¹

Step 2: Calculating Fitness of an Individual. Fitness is a measure of how similar $I_{\text{comp}}(q)$ and $I_{\text{exp}}(q)$ are, and a high fitness implies a close match between $I_{\text{comp}}(q)$ and $I_{\text{exp}}(q)$. For the range of q values explored, we define a sum of (log difference) squared errors between the two functions as

$$\text{sse} = \sum w \left[\log \left(\frac{I_{\text{exp}}(q)}{I_{\text{comp}}(q)} \right) \right]^2 \quad (9)$$

where w is a weighting factor proportional to $\log(\Delta q)$ in order to give similar weights to $I(q)$ values in the range of q values explored. We use log difference squared errors to give similar importance to features at different orders of magnitude. The sum of square errors is minimal when $I_{\text{comp}}(q)$ matches $I_{\text{exp}}(q)$ and large when the two functions differ. To define a fitness function, we rescale sse as

$$\text{Fitness} = X(\text{sse}_{\text{max}} - \text{sse}) + Y \quad (10)$$

where sse_{max} is the maximum value of sse for an individual in the current population. The subtraction of $\text{sse}_{\text{max}} - \text{sse}$ ensures that fitness is maximum when $I_{\text{comp}}(q)$ matches $I_{\text{exp}}(q)$. The parameters X and Y serve the purpose of ensuring that low fitness solutions are not eliminated completely from the population to ensure the GA does not prematurely converge. In Section S.2, we describe in detail how the parameters X and Y are calculated.

Step 3: Selecting Individuals That Go to the Next Generation.

Selection refers to the process of choosing which individuals of the population are preserved in the following generation. The probability that an individual is selected for the next generation is proportional to its fitness values, $P_{\text{selection}} \sim \text{Fitness}$. Selection ensures that individuals with high fitness are more likely to be preserved, while individuals with low fitness are more likely to be discarded but not completely eliminated. Our implementation of GA uses a population of 100 individuals, out of which 99 are selected using the procedure described above, and the remaining one is always the individual with the highest fitness.³⁵

Step 4: Crossover and Mutation.

Crossover is the process in which two selected “parent” individuals are mixed to produce a new “offspring” individual that, in theory,³⁵ combines desirable characteristics from its parents and leads to a higher fitness. A selected individual is chosen for crossover with a probability PC. Once an individual is chosen for crossover, a second “parent” individual is selected using the selection criterion explained above. An “offspring” individual has characteristics (N_{agg} , E_{corona} , and $I_{\text{background}}$) randomly mixed from the two parents. Mutation is the process in which a selected individual is changed randomly to produce a new individual that brings new features to the population that are not present before. A selected individual is chosen for mutation with a probability PM. Once an individual is chosen for mutation, a randomly chosen characteristic (N_{agg} , E_{corona} , or $I_{\text{background}}$) is changed randomly within its range of possible values. While crossover is meant as a mechanism to improve population fitness, mutation is meant to be a process to maintain population diversity to avoid early convergence in local optima. Probabilities PC and PM are, thus, updated throughout a GA run. In Section S.2, we describe in detail the crossover and mutation procedures as well as the choice and update of PC and PM and how N_{agg} , E_{corona} , and $I_{\text{background}}$ are updated.

Step 5: Terminating the Algorithm. The algorithm finishes when the average population fitness and the fitness of the best individual reach a plateau. In this work, the GA is run for a total of 150 generations. All cases studied herein result in converged fitness before the 150 generation mark. D_{core} , D_{micelle} , and N_{agg} are reported as the average and standard deviation from the individual with the highest fitness among 10 independent GA runs.

II.A.3. Micelle Reconstruction. After N_{agg} , D_{core} , and D_{micelle} have been determined by GA, we reconstruct the micelles with molecular detail using CGMD simulations. We take D_{core} , D_{micelle} , and N_{agg} from the individual with highest fitness from each of the 10 independent GA runs; this leads to 10 independent micelle reconstruction results.

Step 1: Determining the Polymer Model. The first step in micelle reconstruction is to determine the parameters of a bead–spring coarse-grained (CG) polymer model to use. In principle one can choose any polymer model representation for these micelle reconstruction simulations based on guidance from extensive previous work on the development of CG models of polymers.^{49,62–65} The choice of the model depends on the desired resolution of micelle features, where more detailed models can be used to determine, for example, the distribution of chemical functional groups within the micelle. Alternatively, the results from CREASE could be used to validate an existing model or to parametrize a desired molecular

model to capture the characteristics of the assembled structure. In this paper, we first determine the degree of coarse graining, i.e., the number of beads of each type, $N_{CG,A}$ and $N_{CG,B}$, and the diameter of those beads, $\sigma_{CG,A}$ and $\sigma_{CG,B}$, that represent the polymer in CGMD simulations. To determine the degree of coarse graining for the solvophobic B block, we match the density of the B polymer in micelle cores, ρ_p , between CGMD simulations and experimental values. The packing fraction of the B polymer in the core, η_{core} , is defined as the volume of CG B beads in the micelle core divided by the micelle core volume and is given by

$$\eta_{core} = \frac{N_{CG,B} N_{agg} \sigma_{CG,B}^3}{D_{core}^3} \quad (11)$$

The density of polymer in micelle cores, ρ_p , is given by

$$\rho_p = \frac{6 N(1 - f_A) N_{agg} M_B}{\pi D_{core}^3} \quad (12)$$

The number of CG B beads, $N_{CG,B}$, at a packing fraction η_{core} corresponding to a B polymer density of ρ_p is obtained by rearranging eqs 11, 12, and 14 (below) as

$$N_{CG,B} = N(1 - f_A) \left(\frac{\pi \rho_p N_{agg} I_{mono,B}^3}{6 M_B \eta_{core}} \right)^{1/2} \quad (13)$$

Equation 13 is comprised of constants that are input to GA (Section II.A.2) and η_{core} . If we know η_{core} we can then fix the degree of coarse graining that matches the polymer density in micelle cores. Our past experience^{50–55} has shown that CG beads that interact with attractive Lennard-Jones (LJ) potentials pack in micelle cores at a packing fraction of $\eta_{core} \sim 0.5$; thus, we use that value to fix the degree of coarse graining. $N_{CG,B}$ is linked to the diameter of CG beads, $\sigma_{CG,B}$, through

$$\sigma_{CG,B} = \frac{N(1 - f_A) I_{mono,B}}{N_{CG,B}} \quad (14)$$

For our case we set the diameter of solvophilic A CG beads to be the same as that of B CG beads (i.e., $\sigma_{CG,A} = \sigma_{CG,B}$), although this can also be the choice of the user or be updated based on the comparison with experiments. From $\sigma_{CG,A}$ we determine the number of CG A beads as

$$N_{CG,A} = \frac{N f_A I_{mono,A}}{\sigma_{CG,A}} \quad (15)$$

Step 2: Reconstructing the Micelle. For simulating the micelle reconstruction, we first create an initial configuration where the N_{agg} (from the individual with the highest fitness in the GA) number of CG polymer chains are placed together side by side in a square array. The choice of the initial configuration is arbitrary and, in our case, is chosen purely for convenience, to ensure that B blocks are localized together and can, in a short time frame, relax away from this configuration while keeping the B beads together in the micelle core. One may choose alternate initial configurations, such as placing all the B blocks of N_{agg} number of chains within a spherical confinement of D_{core} and letting the system relax from that initial configuration under the spherical confinement.

During the relaxation stage of the simulation (see Section II.B.1 for simulation details), the system evolves away from the chosen initial configuration and into spherical micelles with a distinct core and corona. There are a few important checks to do during this relaxation stage to determine whether the reconstruction is successful. The attractive interaction strength between solvophobic B beads, ϵ_{BB} (see Section II.B.1 for model and simulation details), needs to be high enough to prevent the polymer chains from disassembling and dispersing but not too high to prevent them from being kinetically trapped in the user-selected initial configuration. If, during the relaxation stage, the chains disassemble to organize themselves into multiple smaller micelles of aggregation number below N_{agg} and then the value of ϵ_{BB} should be increased. The simulation should be rerun

from the initial configuration. It is possible to predict a value for ϵ_{BB} based on the specific polymer chemistry and solvent/solvent mixture. As an example, we direct the reader to the supplementary sections of our recent work^{52,53} where we connected ϵ_{BB} to the Flory–Huggins χ parameter using the Hildebrand solubility theory. Nonetheless, a discrepancy between D_{core} from the GA and D_{core} from micelle reconstruction indicates that the CG model determination needs to be revisited. In all cases reported here, the resulting D_{core} from micelle reconstruction is comparable to GA results. A discrepancy between $D_{micelle}$ from the GA and $D_{micelle}$ from micelle reconstruction is a result of incorrect assumptions for solvophilic A bead diameter and interactions. If $D_{micelle}$ from micelle reconstruction is smaller than that from GA prediction, we can add a soft repulsive potential between A beads—equivalent to electrostatic repulsion in a charged polymer. If $D_{micelle}$ from micelle reconstruction is greater than that from GA, then a weak attractive potential between A beads can be added—equivalent to the solvophilic block being in a slightly poor solvent. In the results presented herein, there is good agreement between $D_{micelle}$ from the GA and $D_{micelle}$ from micelle reconstruction, and therefore we did not need to change the assumed values for solvophilic bead size or interactions.

Once the micelle reconstruction simulations are validated, we report N_{agg} , D_{core} , and $D_{micelle}$, as well as microscopic information about the micelles, including the radius of gyration of the A block and the spatial arrangement of species A and B monomers within the micelle.

II.B. CGMD Simulation Details. The role of CGMD simulations in this paper is 2-fold. First, we use CGMD simulations for micelle reconstruction to elucidate the chain- and monomer-level structural details of the micelles for the best individuals produced by GA. Second, we use CGMD simulations for the “simulated experiments” where we sample a range of micelle configurations for a variety of polymer designs and generate different kinds of *in silico* experimental $I(q)$ to test the ability of CREASE to predict micelle structures for those *in silico* $I_{exp}(q)$, prior to testing CREASE for *in vitro* SANS-generated $I_{exp}(q)$.

To run these CGMD simulations, we need a CG polymer model. We use CG models identical to the ones used in our recent work,^{50–53,55} where BCPs have a solvophilic A block and a solvophobic B block⁵³ and are modeled as flexible CG bead–spring chains.^{66,67} The neighboring bonded CG beads are connected through a harmonic spring potential with an equilibrium bond distance σ and force constant equal to $50 \text{ kT}/\sigma^2$, where σ is the diameter of CG beads; k is Boltzmann’s constant; and T is the absolute temperature. We model the solvent implicitly and capture the effect of solvent quality on the behavior of the amphiphilic chains through nonbonded interactions. To model the poor solvent quality on the solvophobic B beads, we use an attractive Lennard-Jones (LJ) potential between the B beads with a cutoff distance equal to 2.5σ . The increasing value of the interaction strength of this LJ potential, ϵ_{BB} , also termed solvophobicity, corresponds to increasingly poor solvent quality for the B beads. As the implicit solvent is a good solvent for block A, we model A–A interactions with the repulsive Weeks–Chandler–Andersen (WCA) potential;⁶⁸ WCA potential is a shifted-truncated LJ potential with a cutoff distance of $2^{1/6}\sigma$. As the B–B interactions are the dominant interaction in the solution that drives micellization of the BCPs, we model the nonbonded A–B interaction also with a repulsive-only WCA potential.

For both micelle reconstruction and the simulated experiments, we run molecular dynamics (MD) simulations using the LAMMPS package⁶⁹ (April 2015 version). These simulations are run in periodic cubic simulation boxes in an NVT ensemble with a Langevin thermostat. Much of our simulation protocol is similar to our recent work,^{50–53,55} and we only present the essential details here.

II.B.1. Details for Micelle Reconstruction. The N_{agg} number of CG polymer chains are placed in a cubic simulation box equal to twice the total chain length plus the cutoff range of nonbonded potentials in order to prevent any self-interaction of the chains across the periodic boundaries. As stated in Section II.A.3, the solvophobic components of the chains are packed together in a square array in the initial

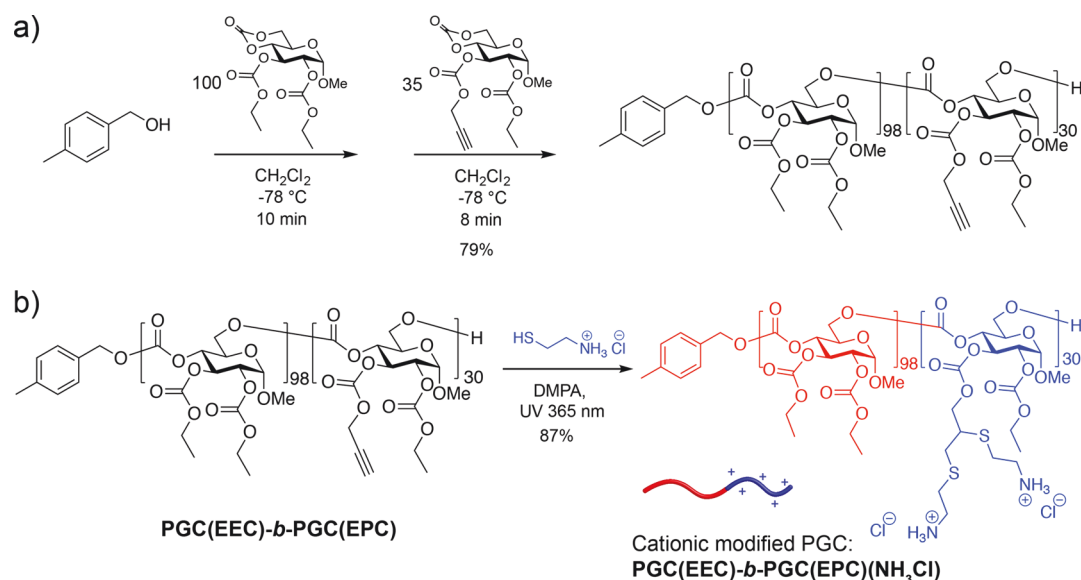


Figure 2. (a) Synthesis of diblock copolymer $\text{PGC(EEC)-}b\text{-PGC(EPC)}$ via sequential ROP. (b) Synthesis of cationic modified amphiphilic diblock copolymers via thiol-yne reaction.

configuration. We evolve the system from this initial configuration by simulating the system for a total of 10 000 000 timesteps at a fixed value of ϵ_{BB} to relax the chains from their initial conformation. Keeping in mind the checks described in Section II.A.3, we run this relaxation stage at $\epsilon_{\text{BB}} = 0.5$; if the system evolves away from the N_{agg} obtained in GA toward smaller micelles, we then increase the ϵ_{BB} in steps of 0.1 until, for a selected ϵ_{BB} , the system maintains the spherical micelle shape at the intended N_{agg} throughout the relaxation from the initial configuration.

As the configurations relax, we calculate the N_{agg} , D_{core} , and D_{micelle} of the micelles. To calculate these parameters, we first identify the cluster(s) or micelle(s) and the chains in each micelle using our protocol described in detail in refs 50–55. For each cluster we also calculate the monomer/bead concentration profiles, $C(r)$, defined as the volume packing fraction of beads at various radial distances starting from the center of mass of each cluster/micelle. In $C(r)$, we consider the solvophilic A beads and solvophobic B beads separately, where the A bead concentration profile indicates the solvophilic micelle corona and the B bead concentration profile indicates the solvophobic micelle core. We use the values of r where the A and B concentration profiles reach 50% of their maximum values to determine D_{micelle} and D_{core} , respectively, as shown in Figure S3. We also calculate the conformations of corona chains by calculating the radius of gyration of the solvophilic A block of each chain, R_g , and report the ensemble average of R_g over all chains in each cluster in multiple configurations.

II.B.2. Details for Simulated Experiments. We simulate the self-assembly of the amphiphilic BCPs in dilute conditions. These simulations are performed on diblock BCPs with a total of 24 CG beads with different compositions: $A_{12}-b-B_{12}$, A_6-b-B_{18} , and $A_{18}-b-B_6$. As these are generic BCPs, the length scale of the simulation is described in units of diameter of 1 CG bead, σ . However, σ can be equated to a real distance if one wishes to model a specific chemistry. For example, the diameter of 1 CG bead, σ , will be equal to 5.04 nm if that CG bead represented approximately 7 repeat units for the PGC polymer synthesized (see section II.C.2). The initial configuration contains 600 polymer chains placed in a cubic simulation box that is larger than the desired occupied polymer volume fraction (η) of 0.025 in order to help the chain relax from their initial configurations and to prevent unrealistic overlap of CG beads. We then linearly reduce each of the simulation box sides over 600 000 timesteps at a temperature of $T^* = 4$ and $\epsilon_{\text{BB}} = 0.055$ until the desired η is attained. For the last step in the equilibration procedure, we set the simulation temperature $T^* = 1$ and further equilibrate for an additional 600 000 time steps.

To drive the self-assembly of the amphiphilic BCPs, we gradually increase the solvophobicity, ϵ_{BB} , in a stagewise manner which is similar to experiments where a good solvent for the solvophobic block is gradually exchanged with a poor solvent for the solvophobic block. We spend 1 000 000 timesteps at each value of ϵ_{BB} , and at the end of each stage we increase ϵ_{BB} in increments of 0.009 until we see no further change in the self-assembly behavior of the BCPs. We extensively tested the amount of time spent at each stage as well as the increase in ϵ_{BB} between stages to ensure that the results are reproducible and are not in trapped states during the simulation procedure. To ensure that we sample uncorrelated simulation configurations, we save the simulation configurations every 100 000 time steps and disregard the results of the first half of the stage.

As done in our previous studies,^{50–55} we analyze the structure of the micelles formed in the simulated experiments by determining the aggregation numbers of the micelles, N_{agg} , the micelle concentration profiles, $C(r)$, the micelle dimensions, D_{core} and D_{micelle} , and the scattering intensity of simulated experiments, $I_{\text{exp}}(q)$. $I_{\text{exp}}(q)$ is determined by

$$I_{\text{exp}}(q) = \sum_{\alpha \in \{A,B\}} \sum_{\beta \in \{A,B\}} b_{\alpha} b_{\beta} F_{b,\alpha}(q) F_{b,\beta}(q) S_{\text{tot}}(q) \quad (16)$$

where b_A , b_B , $F_{b,\alpha}(q)$, and $F_{b,\beta}(q)$ are defined in section II.A.1. $S_{\text{tot}}(q)$ is the structure factor of all beads in the simulation and is calculated by

$$S_{\text{tot}}(q) = 1 + \frac{4\pi\rho_{\text{tot}}}{q} \int r(g_{\text{tot}}(r) - 1) \sin(qr) dr \quad (17)$$

where ρ_{tot} is the total bead number density in the simulation, and $g_{\text{tot}}(r)$ is the radial distribution function of all beads within the simulation.

II.C. Synthesis of Cationic-Modified PGC Block Copolymers.

A cationic-modified diblock BCP of poly(D-glucose carbonate), $\text{PGC(EEC)-}b\text{-PGC(EPC)(NH}_3^+\text{Cl}^-)$, is used in this work as an experimental evaluation of the capabilities of CREASE for sphere-forming BCPs. This PGC polymer is at a three-to-one degree of polymerization ratio of the solvophobic (PGC(EEC))-to-solvophilic ($\text{PGC(EPC)(NH}_3^+\text{Cl}^-)$) block component, with cationic characteristics on the solvophilic block.

II.C.1. Monomer Synthesis. Two different monomers are selected for the construction of the degradable sugar-based functional diblock copolymers. The bicyclic alkyne-functionalized glucose carbonate methyl-2-O-ethoxycarbonyl-3-O-propargyloxycarbonyl-4,6-O-carbon-

yl- α -D-glucopyranoside GC(EPC) with an alkynyl function group provides the capability of various click-type chemistries including azide–alkyne cycloaddition and thiol–yne reaction. The bicyclic nonreactive glucose carbonate monomer methyl-2,3-O-ethoxycarbonyl-4,6-O-carbonyl- α -D-glucopyranoside GC(EEC) is used to serve as the hydrophobic segment. GC(EPC) and GC(EEC) are synthesized following previously reported methods.⁷⁰

II.C.2. Synthesis and Postpolymerization Modification of the Functional, Degradable Diblock Copolymers. The degradable diblock copolymer PGC(EEC)-*b*-PGC(EPC) is synthesized by rapid one-pot sequential organocatalyzed ROP at -78°C in dichloromethane with 1,5,7-triazabicyclo-[4.4.0]dec-5-ene (TBD) as the organocatalyst (Figure 2a). The monomer feed ratio [GC(EEC)]/[GC(EPC)] is set to 3:1 to provide potentially interesting morphologies and sufficient length of the polymer as predicted by computational results.⁵³ The first block is allowed to proceed for 10 min before addition of the second monomer GC(EPC), and the reaction is quenched by addition of an excess amount of acetic acid after 8 min of the subsequent polymerization. Size exclusion chromatography (SEC) of the crude product shows less than 5% remaining monomers, indicating quantitative conversions of both monomers (Figure S4b). The shorter retention time of PGC(EEC)-*b*-PGC(EPC) relative to the first block homopolymer PGC(EEC) confirms the successful chain extension to afford the expected diblock copolymer. SEC traces further reveal monomodal and narrow molecular weight distribution and low dispersity ($\bar{D} < 1.10$), demonstrating the well-defined structure of the diblock copolymer. The degrees of polymerization (DP_n) of each block are calculated from the ^1H NMR spectrum (Figure S4c) acquired after isolation of the polymer by precipitation, by comparing the integration of the methyl proton resonance at 2.24 ppm from the initiator with the integrations of the resonances of the anomeric proton (5.02, 5.06 ppm) and 4-O proton resonances (4.84, 4.92 ppm) in the PGC(EPC) and PGC(EEC), respectively.

The cationic copolymer PGC(EEC)-*b*-PGC(EPC)(NH_3Cl) is obtained by postpolymerization modification of the diblock copolymer PGC(EEC)-*b*-PGC(EPC) with an excess (5 equiv) of cysteamine via photoinitiated thiol–yne click reaction (Figure 2b). The reaction mixture is allowed to stir in *N,N*-dimethylformamide (DMF) under 365 nm in a UV cross-linker at room temperature for 2 h, followed by dialysis against nanopure water (pH = 3 with addition of HCl) at 4°C for 2 days to remove excess thiol and photoinitiator. Finally, the obtained solution is lyophilized to afford fluffy white powders in 80–90% yield. The presence of the thioether side chain group proton resonances in the ^1H NMR spectra (Figure S5) indicates quantitative consumption of the alkyne groups. The appearance of frequencies at $3650\text{--}2300\text{ cm}^{-1}$ in the Fourier-transform infrared spectroscopy (FT-IR) further suggests the introduction of amine groups (Figure S6).

II.D. PGC Micelle Sample Preparation and Experimental Micelle Characterization. The BCPs synthesized following the protocol shown in the previous section are dispersed directly in water, where they self-assemble into micelles. For transmission electron microscopy (TEM) imaging, the PGC is dissolved in pure Milli-Q water. For SANS sample preparation, the PGC is dissolved in deuterated water in order to provide isotopic contrast. For TEM and SANS, the samples are prepared at 0.01, 0.1, and 0.5 wt % block copolymer concentrations.

II.D.1. Transmission Electron Microscopy (TEM). TEM experiments are performed on an FEI TALOS microscope operating at an accelerating voltage of 200 keV, equipped with an FEI Falcon II camera. For TEM grid preparation, a droplet of polymer solution is deposited directly onto a plasma-treated carbon-coated copper grid and dried at room conditions. Before the solution is completely evaporated, a droplet of freshly prepared phosphotungstic acid (2 wt %) is deposited for negative staining, and the excess solution is blotted. For cryogenic-TEM sample preparation, the vitrified grid preparation is done by using the FEI Vitrobot system, an automated plunge vitrification device. A droplet (3 μL) of polymer micelle solution is deposited onto a plasma-treated lacey carbon grid and then

quickly plunged into a liquid ethane reservoir. The grids are then transferred to liquid nitrogen until imaging. During the imaging, the temperature is maintained at -176°C to prevent aqueous solvent crystallization. All images are analyzed with Fiji-ImageJ software.⁷¹ The average size of the particles and the error are calculated from the mean and standard deviation values, respectively, with sample size, $n > \sim 50$ particles.

II.D.2. Small-Angle Neutron Scattering (SANS). SANS experiments are conducted using the NGB7 30m SANS instrument at the Center for Neutron Research (NCNR) in the National Institute for Standards and Technology (NIST, Gaithersburg, MD). Three configurations at 1, 4, and 13 m sample-to-detector distances are used to collect $I(q)$ at a q -range spanning $0.003\text{--}0.5\text{ \AA}^{-1}$ with cell thickness of 1 mm. All experiments are carried out at ambient conditions. All data are corrected for the instrument background, empty cell scattering, and detector efficiency/sensitivity, as well as converted to absolute scale in units of cm^{-1} based on the NIST NCNR data reduction protocol.⁷² The three concentrations are used to find the optimal concentration with minimum micelle–micelle interactions. SANS $I(q)$ is fitted initially using the SASVIEW software⁷³ using the polymer micelle model developed by Pedersen and co-workers.³⁰ This model describes a spherical micelle core with corona chains as noninteracting Gaussian chains attached to the surface of the core. In section S.5 we include the polymer micelle model. Our fit does not include polydispersity effects, and the corona penetration factor is 1.0. This assumes that the distance between the center of the corona and the center of the core is $0.5D_{\text{core}} + R_g$ with no overlapping core/corona region.

The parameters we extract from SASVIEW fits are micelle core diameter (D_{core}), radius of gyration of the corona chains (R_g), and aggregation number (N_{agg}). For the SASVIEW fit, data for $q < 0.02\text{ \AA}^{-1}$ are ignored, so micelle–micelle interactions are not considered.

III. RESULTS AND DISCUSSION

We start by evaluating CREASE for the scattering profiles obtained from the simulated experiments where micelle dimensions, aggregation number, chain conformation, and spatial polymer block distribution are known based on averages collected from the simulated trajectories. As explained in the Methods section, the results from simulated experiments are used to generate a scattering intensity, $I_{\text{exp}}(q)$, that serves as an input to CREASE. Comparing the output of CREASE to the known values of micelle dimensions, aggregation number, chain conformation, and spatial polymer block distribution allows us to establish how well CREASE predicts these known values. After evaluating CREASE against simulated experiments, we test the ability of CREASE with $I_{\text{exp}}(q)$ obtained from experiments, in particular SANS characterization of BCP micelles; we use electron microscopy images to infer micelle shapes which is also an input to CREASE. The output from CREASE is then compared to the results obtained by fitting the spherical polymer micelle model³⁰ to the $I_{\text{exp}}(q)$ and to dimensions of micelles obtained from cast film TEM and cryogenic-TEM analysis. These comparisons establish that CREASE is a viable computational tool to analyze scattering results of BCP solutions to predict the assembled structure without having to fit the $I_{\text{exp}}(q)$ with a known model.

III.A. Evaluation of CREASE Using Scattering Profiles from Simulated Experiments. The CGMD simulation trajectories of three solutions containing $\text{A}_{12}\text{-b-B}_{12}$, $\text{A}_6\text{-b-B}_{18}$, and $\text{A}_{18}\text{-b-B}_6$ ⁵³ BCPs show the formation of spherical micelles of similar D_{micelle} but varying D_{core} and N_{agg} and serve as three different tests of the capabilities of CREASE.

The $\text{A}_{12}\text{-b-B}_{12}$ BCPs at a BCP packing fraction of $\eta = 0.025$ form spherical micelles with aggregation number $N_{\text{agg}} \sim 30$ chains/micelle ($\pm \sim 7$), $D_{\text{core}} \sim 40\text{ nm}$ ($\pm \sim 5\text{ nm}$), and D_{micelle}

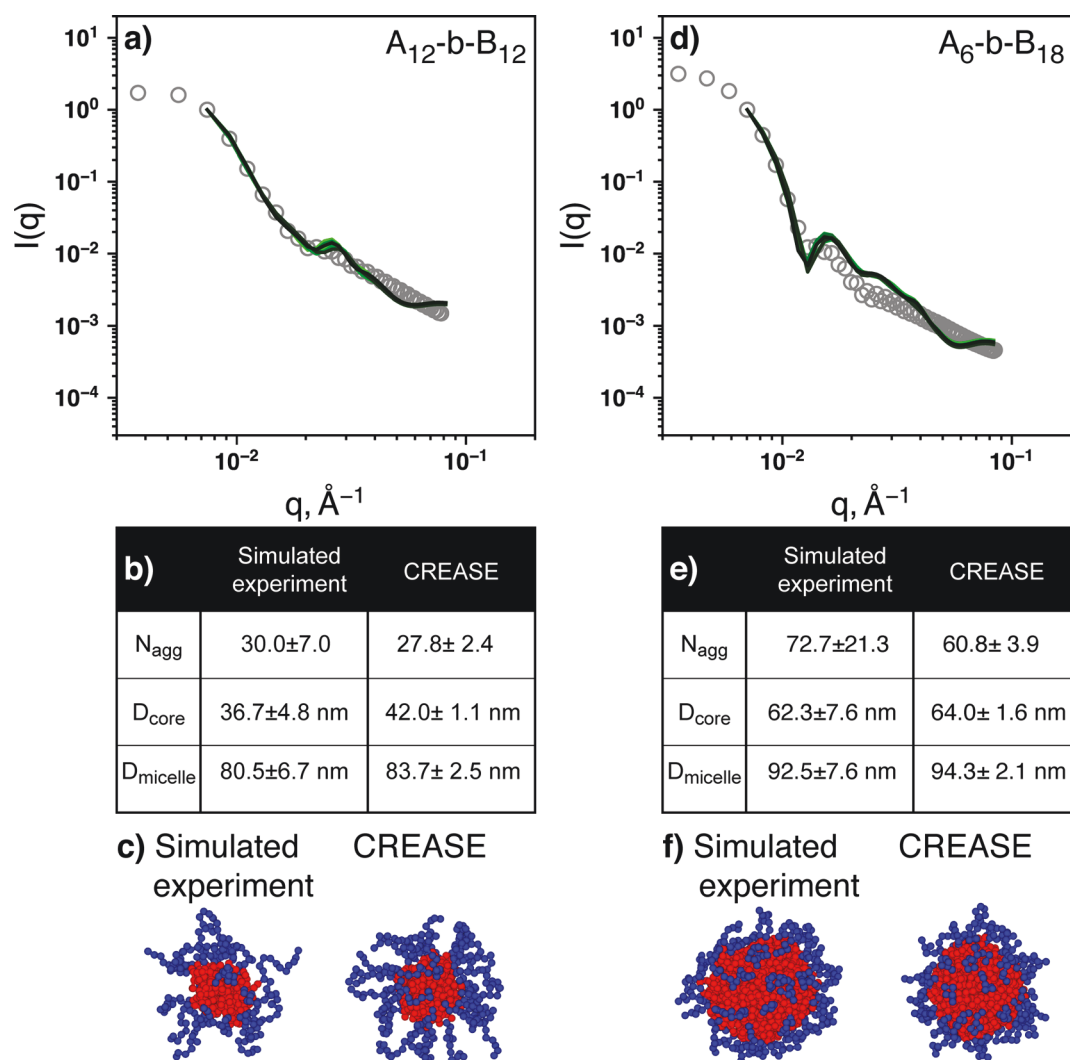


Figure 3. Evaluation of CREASE for scattering intensity results from simulated experiments of $A_{12}\text{-}b\text{-}B_{12}$ BCP (a, b, and c) and $A_6\text{-}b\text{-}B_{18}$ BCP (d, e, and f) in solution. Panels (a) and (d) show scattering intensity from simulated experiment, $I_{\text{exp}}(q)$, in open symbols and from CREASE results, $I_{\text{comp}}(q)$, in lines. Different colored lines represent different independent runs of CREASE. Panels (b) and (e) show aggregation number (N_{agg}), micelle core (D_{core}), and total micelle (D_{micelle}) diameters for the simulated experiment and from CREASE. Reported values for simulated experiments correspond to average and standard deviation from ~ 20 micelles present in the simulation box in 50 independent configurations. Reported values for CREASE correspond to the average and standard deviation from 10 independent runs of the methods. Panels (c) and (f) show renderings of typical micelles from the simulated experiment and CREASE results.

~ 80 nm ($\pm \sim 7$ nm). The $I_{\text{exp}}(q)$ obtained from the CGMD simulations of this system (Figure 3a) is characterized by an initial decrease at low wavevector, q , followed by a mid- q oscillation, followed by further decrease. We carefully choose the range of q values to fit with GA as there are inaccuracies at low values of q due to the finite simulation box size limitations, and we are limited at high q values by the size of the scatterers. The range of $I_{\text{exp}}(q)$ used as an input for CREASE encompasses part of the low- q decrease, the mid- q oscillation, and part of the high- q decrease. The computed scattering profiles from CREASE, $I_{\text{comp}}(q)$, shown in green in Figure 3a agree with $I_{\text{exp}}(q)$ in the range of q values (0.038 to 0.078 $\text{\AA}^{-1}$) where CREASE is run, capturing the mid- q oscillation. The table in Figure 3b shows excellent overall agreement in the values of N_{agg} , D_{core} , and D_{micelle} from the simulated experiment and the corresponding CREASE predictions. CREASE overestimates D_{core} slightly, and this may be because $I_{\text{exp}}(q)$ includes the contribution of a polydisperse group of micelles. CREASE in its current form assumes monodisperse micelles.

Polydispersity is known to shift the mid- q oscillations to lower- q values in dynamic light-scattering experiments.⁷⁴ Analysis of mid- q oscillations that are shifted without accounting for polydispersity leads to biasing the analysis toward larger micelles. Figure 3c shows a typical micelle formed in an MD simulation next to a snapshot of a micelle reconstructed from CREASE, thus serving as a visual confirmation of the predictive power of CREASE.

The $A_6\text{-}b\text{-}B_{18}$ BCPs also form spherical micelles with $N_{\text{agg}} \sim 70$ chains/micelle ($\pm \sim 20$), $D_{\text{core}} \sim 60$ nm ($\pm \sim 8$ nm), and $D_{\text{micelle}} \sim 90$ nm ($\pm \sim 8$ nm). The scattering intensity from the simulated experiments, $I_{\text{exp}}(q)$, is shown in Figure 3d and is characterized by similar features as that for $A_{12}\text{-}b\text{-}B_{12}$, i.e., low- q decrease, mid- q secondary maximum, and high- q decrease. The range of $I_{\text{exp}}(q)$ used as an input for CREASE encompasses, similar to $A_{12}\text{-}b\text{-}B_{12}$, part of the low- q drop, the mid- q oscillation, and part of the high- q drop. CREASE $I_{\text{comp}}(q)$ (solid curves in Figure 3d) agrees with $I_{\text{exp}}(q)$ in the range of q values where CREASE is run, capturing the mid- q oscillation.

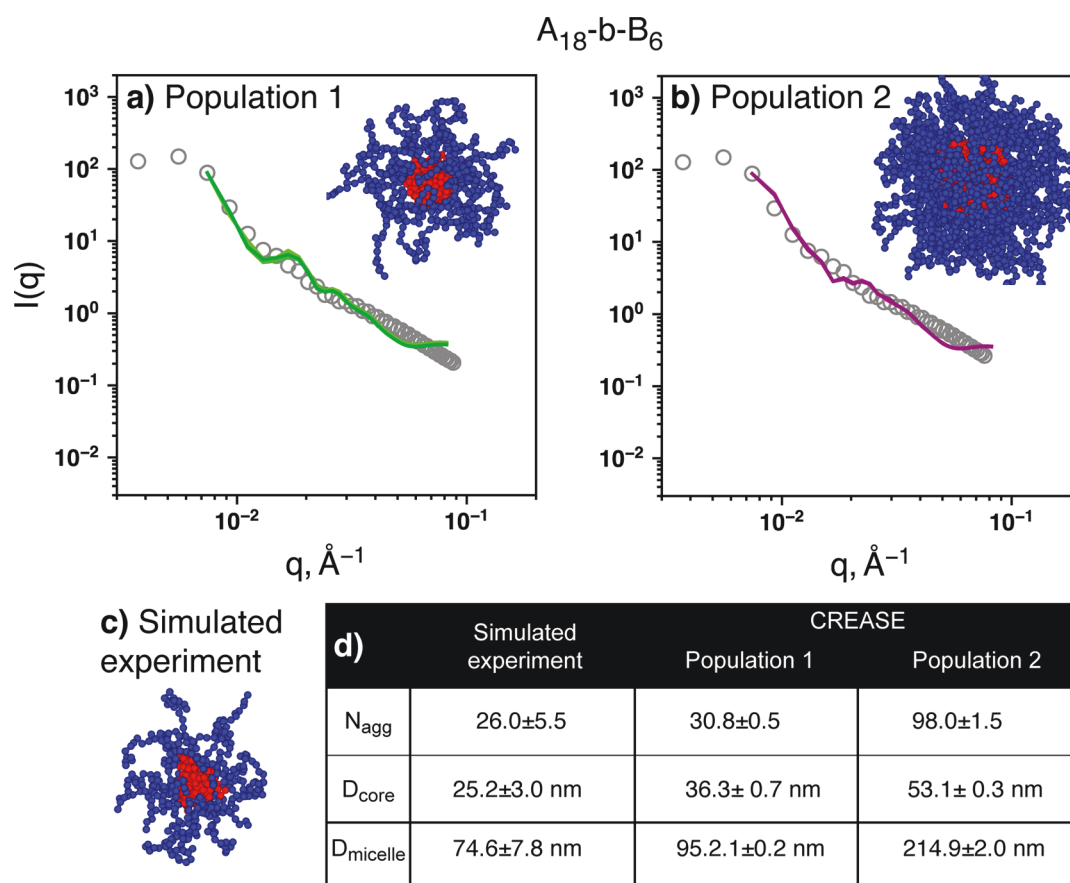


Figure 4. Evaluation of CREASE for scattering intensity results from simulated experiments of $A_{18}\text{-}b\text{-}B_6$ BCP in solution, showing multiple possible populations. Panels (a) and (b) show scattering intensity from the simulated experiment, $I_{\text{exp}}(q)$, in open symbols and from CREASE results, $I_{\text{comp}}(q)$, in lines for population 1 and 2, respectively, of CREASE results. Different colored lines represent different independent runs of CREASE. Insets in panels (a) and (b) show renderings of micelles from CREASE results. Panel (c) shows a rendering of typical micelles from the simulated experiment. Panel (d) shows aggregation number (N_{agg}), micelle core (D_{core}), and total micelle (D_{micelle}) diameters for the simulated experiment and for CREASE results. Reported values for simulated experiments correspond to average and standard deviation from ~ 20 micelles present in the simulation box in 50 independent configurations. Reported values for CREASE correspond to average and standard deviation from 10 independent runs.

The table in Figure 3e shows excellent agreement between D_{core} and D_{micelle} from the simulated experiment and CREASE. The value of CREASE N_{agg} falls within the distribution of N_{agg} values sampled in the simulated experiment. Further visual confirmation of the predictive capability of CREASE is seen in Figure 3f where micelles from simulated experiment and CREASE prediction are shown side-by-side.

The $A_{18}\text{-}b\text{-}B_6$ BCPs also form spherical micelles with $N_{\text{agg}} \sim 25$ chains/micelle, $D_{\text{core}} \sim 25$ nm, and $D_{\text{micelle}} \sim 75$ nm. These $A_{18}\text{-}b\text{-}B_6$ micelles are similar to the $A_6\text{-}b\text{-}B_{18}$ micelles, but the corona is larger and denser in the former case. The $I_{\text{exp}}(q)$ for the $A_{18}\text{-}b\text{-}B_6$ micelles is shown in Figures 4a and 4b; the same plot is shown in two parts to show two different CREASE fits. The range of $I_{\text{exp}}(q)$ used as an input for CREASE encompasses, similar to the case for $A_{12}\text{-}b\text{-}B_{12}$, part of the low- q decrease, the smoothed mid- q oscillation, and part of the high- q decrease. We see that the CREASE $I_{\text{comp}}(q)$ agrees with $I_{\text{exp}}(q)$ in the range of q values where CREASE is run; nonetheless, the location of the mid- q oscillation is not clear, and as a result, CREASE provides two populations with best fitness. Population 1 (Figure 4a) corresponds to micelles with features similar to the simulated experiment. The snapshot of the micelle in Figure 4a from CREASE and the snapshot in Figure 4c for the simulated experiment show that the micelles

are visually similar. The table in Figure 4d shows that micelle $N_{\text{agg}} \sim 25$ for the simulated experiment and $N_{\text{agg}} \sim 30$ for CREASE results are close, but the core and total micelle diameters predicted by CREASE are larger than those of the simulated experiment. Population 2 (Figure 4b and the table in Figure 4d) corresponds to larger micelles predicted from CREASE as compared to those in the simulated experiment. This is a case where our results from simulated experiments guide us in identifying which of the two populations is closer to the correct answer. We also note that the poor prediction from CREASE for $A_{18}\text{-}b\text{-}B_6$ BCPs is likely due to the reduced certainty of the location of the mid- q oscillation that was well resolved for the previous two BCP cases (Figure 3) of $A_{12}\text{-}b\text{-}B_{12}$ and $A_6\text{-}b\text{-}B_{18}$. In similar scenarios in experiments where one may have *in vitro* $I_{\text{exp}}(q)$ with higher uncertainty in one or more q regions in the $I_{\text{exp}}(q)$ profile, to identify/isolate which (if at all) of the populations predicted by CREASE is correct, we need another independent, visual measurement of micelles (e.g., electron microscopy).

III.B. Experimental Characterization of PGC BCP Micelles. The PGC micelle structures in water are first examined using TEM. Both dried cast film and cryogenic-TEM data for PGC at 0.1 wt % show that the resulting structures are spherical micelles, as seen in Figure 5. These micelles each

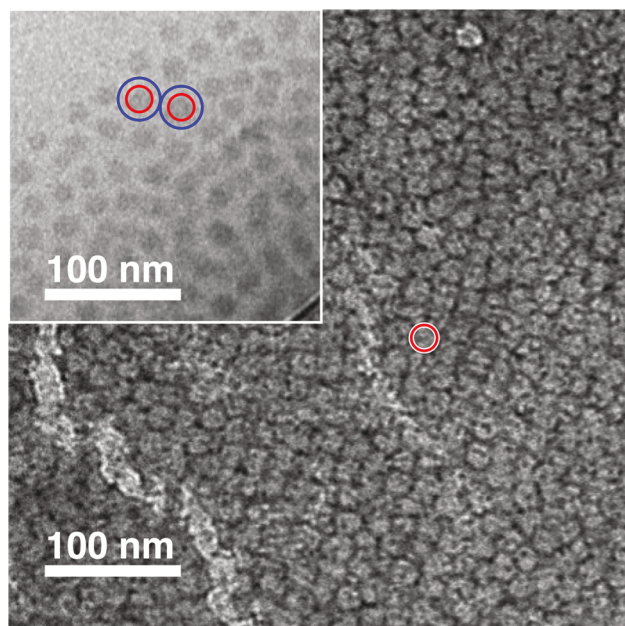


Figure 5. TEM images of 0.1 wt % cationic-modified PGC BCP micelles in water are shown. The main image shows cast film TEM, while the inset shows a cryogenic-TEM image. In the main image, the red line encircles a micelle core with $D_{\text{core}} \sim 20$ nm. In the inset, red lines encircle two adjacent micelle cores, and blue lines encircle the probable corona chain boundaries with $D_{\text{micelle}} \sim 30$ nm.

have a hydrophobic PGC(EEC) core with charged PGC-(EPC)(NH_3^+) in the corona. The dark spheres in the cryo-TEM micrograph represent primarily the denser hydrophobic cores of the micelles. The average diameter of these cores, from a sample of ~ 50 micelles in the field of view, is ~ 20.0 nm with a standard deviation of 0.2 nm. Figure 5 highlights sample micelle cores encircled in red for both dried cast film and cryogenic-TEM. Although the contrast of the corona blocks is too low to be visible in the TEM images, the distances between each core of close-packed regions in cryogenic-TEM data are estimated to provide an approximate total micelle diameter, D_{micelle} . The inset in Figure 5 shows two micelles with the measured D_{core} and D_{micelle} highlighted side-by-side to provide a visual representation of micelle dimensions in scale. Assuming that corona chains of neighboring micelles are not overlapping, these micelles have a $D_{\text{micelle}} \sim 29.2$ nm and a corona shell thickness, $2R_g$, of approximately 4.6 nm. The TEM-measured values of D_{core} , D_{micelle} , and corona chain R_g for 0.01, 0.1, and 0.5 wt % PGC BCP solutions are presented in the next section. Figure S7 shows TEM images for 0.01 and 0.5 wt % PGC micelles.

TEM data provide important real space information on the micelle structures but are limited to the sample size captured in a field of view, as seen in Figure 5. SANS, on the other hand, provides reciprocal space information from a much larger, bulk-like sample size and can access larger length scale features. The SANS results of a concentration series of 0.01, 0.1, and 0.5 wt % PGC BCP in D_2O are shown in Figure 6 with scattering intensity $I_{\text{exp}}(q)$ plotted as a function of wavevector, q . For all three concentrations, a prominent oscillation at the mid- q region indicates a regular nanostructure of the micelles.⁷⁵ The PGC block copolymers are unique in that the entire backbone is the same poly(D-glucose carbonate) chain with the hydrophobic core and hydrophilic corona blocks created

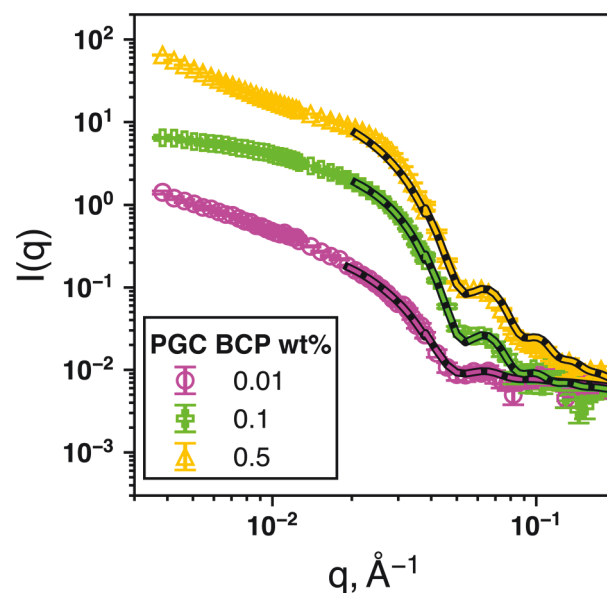


Figure 6. SANS scattering intensities for PGC block copolymers in D_2O for 0.01 (pink), 0.1 (green), and 0.5 (orange) wt %. Symbols show SANS data with standard error, and dashed lines show SASVIEW fits.

with different side chain chemical functionality (Figure 2b). Additionally, the reported persistence length, L_p , for typical neutral glucose backbone-based polymers (i.e., cellulose derivatives) can range from several to tens of nanometers.⁷⁶ This chain stiffness would more closely exhibit rigid rod-like chain behavior rather than flexible, vinyl polymer behavior ($L_p \sim 1$ nm). As mentioned in the Introduction, the primary strength of CREASE is its ability to analyze scattering profiles without a scattering model and provide chain- and monomer-level structural information that is otherwise difficult to obtain from scattering and microscopy alone. This is particularly important for the assembly of PGC BCPs since there is no established model that perfectly describes the packing of such stiff chains into a micelle-like nanoparticle with charged, stiff corona chains. To highlight the importance of the CREASE method, we use the most appropriate model available, the polymer micelle model³⁰ in SASVIEW,⁷³ to fit the experimental SANS data with the best fit lines shown as solid lines in Figure 6. SASVIEW fit results for the micelle core diameter, $D_{\text{core}} \sim 19$ nm, are consistent with the cores observed in the TEM data. We expect micelle dimensions from SASVIEW fits of SANS data not to change with changing polymer concentration because the critical micelle concentration of block copolymers is typically extremely low,^{77–79} and PGC(EEC) cores are likely glassy as the pure block copolymer glass transition temperature, T_g , is >100 °C.⁷⁰ The observed consistency between micelle core approximate size in SANS and cryogenic-TEM makes sense since the core provides the most contrast in both scattering and microscopy. However, the polymer micelle model is meant for flexible chain amphiphilic block copolymers with no significant chain stiffness or long-range, repulsive interactions between corona chains.^{30,31} Therefore, the use of an inappropriate model provides fitted N_{agg} and corona R_g values with no consistency within the range of concentrations as well as physically unrealistic results. Particularly, N_{agg} and corona R_g show that high variability (Table 2) with SASVIEW fits of the corona chain R_g is much

larger than the cryogenic-TEM measurements shown in Figure 5. Furthermore, we can compare the fitted N_{agg} from SASVIEW to an estimated N_{agg} value of approximately 69 calculated using the approximate volume of the micelle core and the known density of hydrophobic PGC chains in water (assuming there is no water present in the hydrophobic core). The reported SASVIEW N_{agg} fit values range from 10 to 30, which are significantly lower than the calculated N_{agg} of 69 using the known PGC density and micelle core diameter. For the N_{agg} from the SASVIEW fit to match the N_{agg} from CREASE, the scattering length densities (β) of the core and corona would have to be extremely large, beyond what is physically possible. If we force both the β_{core} and β_{corona} to be within reasonable values and get the SASVIEW N_{agg} close to the CREASE N_{agg} , the overall SASVIEW fit to the scattering data is poor (data not shown). Since most fit parameters have a narrow range of sensible values (concentration, D_{micelle} , β_{core} , β_{corona} , and β_{solvent}), the reported SASVIEW N_{agg} is from the result that gave us the best data fit. The SASVIEW fit results show a clear need for a model-independent method like CREASE for describing the nanostructure giving rise to the SANS data.

III.C. CREASE Characterization of PGC BCP Micelles from SANS Scattering Intensities. In this section, we use CREASE to characterize the PGC micelles analyzed in the previous section. We start by listing the input parameters to CREASE from known information from polymer synthesis and then proceed to show the output from CREASE and comparison to experimental characterization.

As stated in the Methods section, the input parameters to CREASE are the $I_{\text{exp}}(q)$, taken here from SANS experiments shown in Figure 6, polymer design, density of the solvophobic block (PGC(EEC)), and shape (spherical). The range of wavevector, q , values we analyze with CREASE is 0.038–0.078 Å^{−1} and corresponds to the mid- q oscillation that characterizes the spherical shape. Table 1 summarizes relevant quantities we use in CREASE runs for PGC BCPs.

CREASE is run ten independent times to determine N_{agg} , D_{core} , D_{micelle} , and corona chain R_g . Reported values represent

Table 1. Input Parameters for CREASE Implementation for PGC Block Copolymers Relevant to Polymer Design and Polymer Density

parameter	value
molar mass of solvophobic block, PGC(EEC) (g/mol)	364
molar mass of solvophilic block, PGC(EPC)(NH ₃ Cl) (g/mol) ^a	601
contour length of solvophobic repeat unit, $l_{\text{mono,B}}$ (nm)	0.72
contour length of solvophilic repeat unit, $l_{\text{mono,A}}$ (nm)	0.72
polymer composition, f_A ^b	0.234 (0.25)
total number of repeat units, N	128
solvophobic block density, ρ_p (g/mL) ^c	1.13 (1.25)
number of point scatterers per chain used in GA	$N_{\text{GA,A}} = 2$, $N_{\text{GA,B}} = 6$
number of CG beads per chain used in micelle reconstruction	$N_{\text{CG,A}} = 28$, $N_{\text{CG,B}} = 84$

^aThe molar mass of the solvophilic block, PGC(EPC)(NH₃Cl), includes the chlorine counterion. CREASE calculations are not affected by the addition of the counterion molar mass. ^bCopolymer composition is rounded up to 0.25 in CREASE calculations to yield integer number of beads. ^cSolvophobic block density increases to value in parentheses due to the rounding of f_A .

the average and standard deviation of those ten runs. The variability of CREASE results in all reported quantities is below 4% (standard deviation/average) which points to a high precision within independent runs. In section S.7, Supporting Figures S8–S10, we show CREASE fits to the scattering intensity where all individuals show low variability in the way they fit SANS data within the fit range. In section III.A, we discuss the possibility of multiple populations exhibiting high fitness, especially when the mid- q oscillation is not well resolved. SANS scattering intensities for PGC block copolymers do not appear to have that issue, which explains the low variability in CREASE fit parameters and scattering intensity fit curves. CREASE predictions of micelle characteristics do not change with increasing PGC concentration, consistent with our expectation given the low critical micelle concentration.

Table 2 shows a summary of CREASE results for PGC block copolymer micelles, alongside TEM and SASVIEW fit parameters. Overall, CREASE predictions of micelle dimensions are in good agreement with independently measured TEM results. The micelle core diameter, $D_{\text{core}} \sim 20$ nm, is consistently found in all three techniques. Total micelle diameter, D_{micelle} from SASVIEW fits (~ 70 nm) is different from that seen in cryogenic-TEM and CREASE (~ 30 nm). As we discussed in the previous sections, the assumptions behind the SASVIEW fit include that the corona chains are Gaussian, which ignores the excluded volume and possible semiflexible or rigid rod behavior of the sugar-derived polymer. These assumptions likely explain the difference between measurements. While the polymer micelle model fit provides N_{agg} that varies between different sample concentrations and is too low to be physically possible, CREASE provides N_{agg} estimates that seem physically appropriate. As previously discussed, if one uses the known solvophobic PGC polymer density, one can estimate that N_{agg} should be ~ 69 which is much closer to the CREASE predictions. Overall, D_{core} values are consistent throughout independent measurements; the cryogenic-TEM micelle observations and CREASE predictions agree well on the total micelle dimensions; and the N_{agg} from CREASE matches closely with the estimate based on core dimensions and solvophobic polymer density. This combination of results supports the accuracy and feasibility of the CREASE method in describing the physical nanostructure that produces the experimental SANS data.

Our results so far have focused on comparing CREASE predictions with micelle-level quantities from the standard spherical polymer micelle model and TEM measurements with overall good results. CREASE also provides additional chain-level detail that is difficult to obtain experimentally. As an example, we show in Figure 7 monomer concentration profiles, $C(r)$, as functions of distance from micelle center of mass, r , for micelles formed from PGC BCPs at 0.1 wt %. Figures S11–S12 show $C(r)$ for other experimental conditions. The kind of information provided by micelle reconstruction, including but not limited to $C(r)$, is useful in many BCP micelle applications, e.g., location and accessibility of functional groups in corona chain blocks.

IV. CONCLUSIONS

We present a computational approach for analyzing scattering profiles from experiments to elucidate both assembled micelle structure as well as chain- and monomer-level structural information. As a proof-of-concept, we analyze the structure of spherical block copolymer micelles where (1) the assembled-

Table 2. Summary of CREASE and Experimental Characterization Results for PGC BCP Micelles

PGC wt %		CREASE ^a	SASVIEW ^b	dried cast TEM ^c	cryogenic-TEM ^d
0.01	N_{agg} , aggregation number	88.7 ± 2.0	7 ± 1 (10)		
	D_{core} , core diameter (nm)	20.0 ± 0.2	15.6 ± 2.6 (18.6)	19.4 ± 0.2	
	$D_{micelle}$, micelle diameter (nm)	32.4 ± 0.4	74.8 ± 9.56 (65.8)		
	R_g of corona chain (nm)	3.0 ± 0.0	14.8 ± 2.3 (11.8)		
0.1	N_{agg} , aggregation number	87.7 ± 2.9	14 ± 2 (16)		
	D_{core} , core diameter (nm)	19.9 ± 0.2	18.2 ± 0.1 (18.6)	19.9 ± 0.2	20.0 ± 0.2
	$D_{micelle}$, micelle diameter (nm)	32.2 ± 0.4	73.4 ± 5.6 (59.4)		29.2 ± 0.4
	R_g of corona chain (nm)	3.0 ± 0.0	13.8 ± 1.4 (10.2)		
0.5	N_{agg} , aggregation number	81.4 ± 0.9	23 ± 5 (32)		
	D_{core} , core diameter (nm)	19.5 ± 0.2	18.0 ± 0.6 (19.4)	19.7 ± 0.3	
	$D_{micelle}$, micelle diameter (nm)	31.5 ± 0.2	34.6 ± 4.63 (32.6)		
	R_g of corona chain (nm)	3.0 ± 0.0	4.14 ± 1.15 (3.3)		

^aCREASE results reported correspond to average ± standard deviation from 10 independent runs. ^bSASVIEW results reported from fits to the polymer micelle model correspond to average ± standard deviation from 5 independent runs of the software. Values in parentheses correspond to the best fit. ^cDried cast TEM results from image analysis correspond to average ± standard deviation from ~100 micelles in the field of view. ^dCryogenic-TEM results from image analysis correspond to average ± standard deviation from ~40 micelles in the field of view.

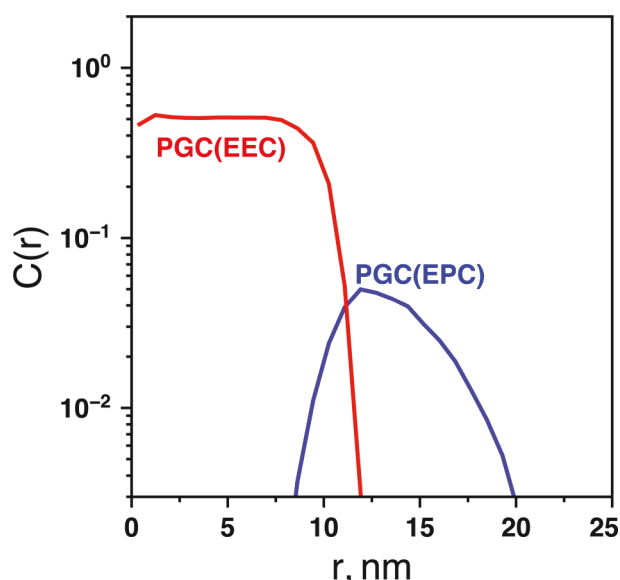


Figure 7. Monomer concentration profile, $C(r)$, as a function of distance from the micelle center of mass, r , for PGC at 0.1 wt %. The red line denotes the solvophobic PGC(EEC), while the blue line denotes the solvophilic PGC(EPC) concentration profiles.

state information includes the micelle core and total diameter and aggregation number and (2) the chain-level information includes chain conformations and spatial arrangement of copolymer blocks within a micelle. This method, CREASE (computational reverse-engineering analysis for scattering experiments), expands the capabilities of traditional analysis techniques for scattering experiments, that typically use approximate models that tie the structure to the scattering intensity, by using assembled-structure information as an input to molecular simulations where features not easily seen experimentally are extracted.

We evaluate CREASE with simulated experiments of spherical micelles formed from block copolymers that we studied in a previous work.⁵³ In our tests we find, in general, comparable values for micelle dimensions between simulated experiments and the results of CREASE. An important exception is in systems with long corona chains, where CREASE finds two populations of solutions that yield good fits

to the simulated experiment scattering intensity. One population yields micelles larger than the input simulated experiment, while the other population yields micelles comparable to the input simulated experiment. We propose a way to unequivocally discern which population to include in the results by realizing that scattering experiments can be made alongside real-space imaging, such as electron microscopy.

After the validation against simulated experiments, CREASE is applied to SANS measurements of spherical-micelle-forming PGC block copolymers, which represent next-generation renewable source-based materials.⁷⁰ Microscopy characterization of the spherical micelles formed by PGC block copolymers agrees well with CREASE results. Fits to SANS data using well-known models^{30,31} that are physically inappropriate for the unique block copolymers reveal the importance of the CREASE method for analysis of scattering experiments of polymer structures.

Future work on this topic includes the extension of CREASE to different micellar shapes and acceleration and streamlining of the algorithm to provide fast results. Given the small number of parameters and assumptions within the framework, along with the information that this method elucidates, which is difficult to extract experimentally, CREASE is a promising technique for characterization of a broad range of assembled polymer materials.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b08028.

Genetic algorithm details; characterization of synthesized PGC; Pedersen and co-workers' polymer micelle model; additional characterization of assembled PGC; additional CREASE results (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*wooley@chem.tamu.edu

*pochan@udel.edu

*arthij@udel.edu

ORCID

Daniel J. Beltran-Villegas: 0000-0003-1667-3181

Arthi Jayaraman: 0000-0002-5295-4581

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge financial support from the National Science Foundation under grant numbers NSF DMREF-1629156 and DMREF-1629094. We also acknowledge the support for the simulation work through the use of information technology resources at the University of Delaware, specifically the Farber high-performance computing resources. Access to NGB-30m SANS instrument was provided by the Center for High Resolution Neutron Scattering, a partnership between the National Institute of Standards and Technology and the National Science Foundation under Agreement No. DMR-1508249. We acknowledge the support of the National Institute of Standards and Technology, U.S. Department of Commerce, in providing the neutron research facilities used in this work. Funding in support of the SANS measurements was provided under cooperative agreements 70NANB12H239 and 70NANB17H302 from NIST, U.S. Department of Commerce. We also acknowledge the support of the National Institute of Standards and Technology, U.S. Department of Commerce, in providing the neutron research facilities used in this work. Neutron facilities are also supported in part by the National Science Foundation under Agreement No. DMR-0944772. Support is also acknowledged, gratefully, from the Welch Foundation as the W. T. Doherty-Welch Chair in Chemistry at Texas A&M University (A-0001).

REFERENCES

- (1) Bates, F. S. Polymer-polymer phase behavior. *Science* **1991**, *251* (4996), 898–905.
- (2) Bates, F. S.; Fredrickson, G. H. Block copolymer thermodynamics: theory and experiment. *Annu. Rev. Phys. Chem.* **1990**, *41* (1), 525–557.
- (3) Riess, G. Micellization of block copolymers. *Prog. Polym. Sci.* **2003**, *28* (7), 1107–1170.
- (4) Mai, Y.; Eisenberg, A. Self-assembly of block copolymers. *Chem. Soc. Rev.* **2012**, *41* (18), 5969–5985.
- (5) Cho, H. K.; Cheong, I. W.; Lee, J. M.; Kim, J. H. Polymeric nanoparticles, micelles and polymersomes from amphiphilic block copolymer. *Korean J. Chem. Eng.* **2010**, *27* (3), 731–740.
- (6) Letchford, K.; Burt, H. A review of the formation and classification of amphiphilic block copolymer nanoparticulate structures: micelles, nanospheres, nanocapsules and polymersomes. *Eur. J. Pharm. Biopharm.* **2007**, *65* (3), 259–269.
- (7) Tritschler, U.; Pearce, S.; Gwyther, J.; Whittell, G. R.; Manners, I. 50th anniversary perspective: Functional nanoparticles from the solution self-assembly of block copolymers. *Macromolecules* **2017**, *50* (9), 3439–3463.
- (8) Cabral, H.; Miyata, K.; Osada, K.; Kataoka, K. Block copolymer micelles in nanomedicine applications. *Chem. Rev.* **2018**, *118* (14), 6844–6892.
- (9) Madsen, J.; Armes, S. P. (Meth) acrylic stimulus-responsive block copolymer hydrogels. *Soft Matter* **2012**, *8* (3), 592–605.
- (10) Pavia-Sanders, A.; Zhang, S.; Flores, J. A.; Sanders, J. E.; Raymond, J. E.; Wooley, K. L. Robust magnetic/polymer hybrid nanoparticles designed for crude oil entrapment and recovery in aqueous environments. *ACS Nano* **2013**, *7* (9), 7552–7561.
- (11) Liang, Y.; Wang, H.; Yuan, S.; Lee, Y.; Gan, L.; Yu, L. Conjugated block copolymers and co-oligomers: from supramolecular assembly to molecular electronics. *J. Mater. Chem.* **2007**, *17* (21), 2183–2194.
- (12) Zhang, Y.; Vallin, J. R.; Sahoo, J. K.; Gao, F.; Boudouris, B. W.; Webber, M. J.; Phillip, W. A. High-Affinity Detection and Capture of Heavy Metal Contaminants using Block Polymer Composite Membranes. *ACS Cent. Sci.* **2018**, *4* (12), 1697–1707.
- (13) Cho, Y.; Sundaram, H. S.; Weinman, C. J.; Paik, M. Y.; Dimitriou, M. D.; Finlay, J. A.; Callow, M. E.; Callow, J. A.; Kramer, E. J.; Ober, C. K. Triblock copolymers with grafted fluorine-free, amphiphilic, non-ionic side chains for antifouling and fouling-release applications. *Macromolecules* **2011**, *44* (12), 4783–4792.
- (14) Reuther, J. F.; Siriwardane, D. A.; Campos, R.; Novak, B. M. Solvent tunable self-assembly of amphiphilic rod-coil block copolymers with chiral, helical polycarbodiimide segments: polymeric nanostructures with variable shapes and sizes. *Macromolecules* **2015**, *48* (19), 6890–6899.
- (15) Gao, Y.; Qiu, H.; Zhou, H.; Li, X.; Harniman, R.; Winnik, M. A.; Manners, I. Crystallization-driven solution self-assembly of block copolymers with a photocleavable junction. *J. Am. Chem. Soc.* **2015**, *137* (6), 2203–2206.
- (16) Greenall, M. J.; Schuetz, P.; Fuzeland, S.; Atkins, D.; Buzza, D. M. A.; Butler, M. F.; McLeish, T. C. Controlling the self-assembly of binary copolymer mixtures in solution through molecular architecture. *Macromolecules* **2011**, *44* (13), 5510–5519.
- (17) Figg, C. A.; Carmean, R. N.; Bentz, K. C.; Mukherjee, S.; Savin, D. A.; Sumerlin, B. S. Tuning Hydrophobicity To Program Block Copolymer Assemblies from the Inside Out. *Macromolecules* **2017**, *50* (3), 935–943.
- (18) Zhang, L.; Eisenberg, A. Multiple morphologies and characteristics of “crew-cut” micelle-like aggregates of polystyrene-*b*-poly (acrylic acid) diblock copolymers in aqueous solutions. *J. Am. Chem. Soc.* **1996**, *118* (13), 3168–3181.
- (19) Song, Y.; Chen, Y.; Su, L.; Li, R.; Letteri, R. A.; Wooley, K. L. Crystallization-driven assembly of fully degradable, natural product-based poly (l-lactide)-block-poly (α -D-glucose carbonate) s in aqueous solution. *Polymer* **2017**, *122*, 270–279.
- (20) Zhu, J.; Jiang, W. Self-assembly of ABC triblock copolymer into giant segmented wormlike micelles in dilute solution. *Macromolecules* **2005**, *38* (22), 9315–9323.
- (21) Zhu, J.; Jiang, Y.; Liang, H.; Jiang, W. Self-assembly of ABA amphiphilic triblock copolymers into vesicles in dilute solution. *J. Phys. Chem. B* **2005**, *109* (18), 8619–8625.
- (22) Cooksey, T. J.; Singh, A.; Le, K. M.; Wang, S.; Kelley, E. G.; He, L.; Vajjala Kesava, S.; Gomez, E. D.; Kidd, B. E.; Madsen, L. A. Tuning Biocompatible Block Copolymer Micelles by Varying Solvent Composition: Core/Corona Structure and Solvent Uptake. *Macromolecules* **2017**, *50* (11), 4322–4334.
- (23) Choucair, A.; Eisenberg, A. Control of amphiphilic block copolymer morphologies using solution conditions. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2003**, *10* (1), 37–44.
- (24) Sprouse, D.; Jiang, Y.; Laaser, J. E.; Lodge, T. P.; Reineke, T. M. Tuning cationic block copolymer micelle size by pH and ionic strength. *Biomacromolecules* **2016**, *17* (9), 2849–2859.
- (25) Shen, H.; Zhang, L.; Eisenberg, A. Multiple pH-induced morphological changes in aggregates of polystyrene-block-poly (4-vinylpyridine) in DMF/H₂O mixtures. *J. Am. Chem. Soc.* **1999**, *121* (12), 2728–2740.
- (26) Zhang, L.; Yu, K.; Eisenberg, A. Ion-induced morphological changes in “crew-cut” aggregates of amphiphilic block copolymers. *Science* **1996**, *272* (5269), 1777–1779.
- (27) Cui, H.; Chen, Z.; Zhong, S.; Wooley, K. L.; Pochan, D. J. Block copolymer assembly via kinetic control. *Science* **2007**, *317* (5838), 647–650.
- (28) Bang, J.; Viswanathan, K.; Lodge, T. P.; Park, M. J.; Char, K. Temperature-dependent micellar structures in poly (styrene-*b*-isoprene) diblock copolymer solutions near the critical micelle temperature. *J. Chem. Phys.* **2004**, *121* (22), 11489–11500.
- (29) 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.

- (30) Pedersen, J. S.; Gerstenberg, M. C. Scattering form factor of block copolymer micelles. *Macromolecules* **1996**, *29* (4), 1363–1365.
- (31) Svaneborg, C.; Pedersen, J. S. A Monte Carlo study on the effect of excluded volume interactions on the scattering from block copolymer micelles. *J. Chem. Phys.* **2000**, *112* (21), 9661–9670.
- (32) Heo, T.-Y.; Kim, I.; Chen, L.; Lee, E.; Lee, S.; Choi, S.-H. Effect of Ionic Group on the Complex Coacervate Core Micelle Structure. *Polymers* **2019**, *11* (3), 455.
- (33) Borbély, S. Aggregate structure in aqueous solutions of Brij-35 nonionic surfactant studied by small-angle neutron scattering. *Langmuir* **2000**, *16* (13), 5540–5545.
- (34) De Jong, K. A. Analysis of the behavior of a class of genetic adaptive systems. *Technical Report No. 185*, 1975.
- (35) Eiben, A. E.; Smith, J. E. *Introduction to evolutionary computing*; Springer: 2003; Vol. 53.
- (36) Chakraborti, N. Genetic algorithms in materials design and processing. *Int. Mater. Rev.* **2004**, *49* (3–4), 246–260.
- (37) Le, T. C.; Winkler, D. A. Discovery and optimization of materials using evolutionary approaches. *Chem. Rev.* **2016**, *116* (10), 6107–6132.
- (38) Arora, V.; Bakhshi, A. Molecular designing of novel ternary copolymers of donor–acceptor polymers using genetic algorithm. *Chem. Phys.* **2010**, *373* (3), 307–312.
- (39) Mitra, K. Genetic algorithms in polymeric material production, design, processing and other applications: a review. *Int. Mater. Rev.* **2008**, *53* (5), 275–297.
- (40) Kasat, R. B.; Ray, A. K.; Gupta, S. K. Applications of genetic algorithm in polymer science and engineering. *Mater. Manuf. Processes* **2003**, *18* (3), 523–532.
- (41) Jaeger, H. M.; de Pablo, J. J. Perspective: Evolutionary design of granular media and block copolymer patterns. *APL Mater.* **2016**, *4* (5), 053209.
- (42) Khaira, G. S.; Qin, J.; Garner, G. P.; Xiong, S.; Wan, L.; Ruiz, R.; Jaeger, H. M.; Nealey, P. F.; De Pablo, J. J. Evolutionary optimization of directed self-assembly of triblock copolymers on chemically patterned substrates. *ACS Macro Lett.* **2014**, *3* (8), 747–752.
- (43) Srinivasan, B.; Vo, T.; Zhang, Y.; Gang, O.; Kumar, S.; Venkatasubramanian, V. Designing DNA-grafted particles that self-assemble into desired crystalline structures using the genetic algorithm. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (46), 18431–18435.
- (44) Chua, A. L.-S.; Benedek, N. A.; Chen, L.; Finnis, M. W.; Sutton, A. P. A genetic algorithm for predicting the structures of interfaces in multicomponent systems. *Nat. Mater.* **2010**, *9* (5), 418.
- (45) Fornleitner, J.; Verso, F. L.; Kahl, G.; Likos, C. N. Genetic algorithms predict formation of exotic ordered configurations for two-component dipolar monolayers. *Soft Matter* **2008**, *4* (3), 480–484.
- (46) Deaven, D. M.; Ho, K.-M. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **1995**, *75* (2), 288.
- (47) Kanters, R. P.; Donald, K. J. CLUSTER: Searching for unique low energy minima of structures using a novel implementation of a genetic algorithm. *J. Chem. Theory Comput.* **2014**, *10* (12), 5729–5737.
- (48) Zhang, Q.; Lin, J.; Wang, L.; Xu, Z. Theoretical modeling and simulations of self-assembly of copolymers in solution. *Prog. Polym. Sci.* **2017**, *75*, 1–30.
- (49) Gartner III, T. E.; Jayaraman, A. Modeling and simulations of polymers: A Roadmap. *Macromolecules* **2019**, *52* (3), 755–786.
- (50) Lyubimov, I.; Beltran-Villegas, D. J.; Jayaraman, A. PRISM theory study of amphiphilic block copolymer solutions with varying copolymer sequence and composition. *Macromolecules* **2017**, *50* (18), 7419–7431.
- (51) Wessels, M. G.; Jayaraman, A. Molecular dynamics simulation study of linear, bottlebrush, and star-like amphiphilic block polymer assembly in solution. *Soft Matter* **2019**, *15* (19), 3987–3998.
- (52) Beltran-Villegas, D. J.; Lyubimov, I.; Jayaraman, A. Molecular dynamics simulations and PRISM theory study of solutions of nanoparticles and triblock copolymers with solvophobic end blocks. *Molecular Systems Design & Engineering* **2018**, *3* (3), 453–472.
- (53) Beltran-Villegas, D. J.; Jayaraman, A. Assembly of Amphiphilic Block Copolymers and Nanoparticles in Solution: Coarse-Grained Molecular Simulation Study. *J. Chem. Eng. Data* **2018**, *63* (7), 2351–2367.
- (54) Dong, M.; Wessels, M. G.; Lee, J. Y.; Su, L.; Wang, H.; Letteri, R. A.; Song, Y.; Lin, Y.-N.; Chen, Y.; Li, R.; et al. Experiments and Simulations of Complex Sugar-Based Coil-Brush Block Polymer Nanoassemblies in Aqueous Solution. *ACS Nano* **2019**, *13* (5), 5147–5162.
- (55) Lyubimov, I.; Wessels, M. G.; Jayaraman, A. Molecular Dynamics Simulation and PRISM Theory Study of Assembly in Solutions of Amphiphilic Bottlebrush Block Copolymers. *Macromolecules* **2018**, *51* (19), 7586–7599.
- (56) Sivia, D. S. *Elementary scattering theory: for X-ray and neutron users*; Oxford University Press: 2011.
- (57) Michalski, R. S. Learnable evolution model: Evolutionary processes guided by machine learning. *Machine learning* **2000**, *38* (1–2), 9–40.
- (58) Patra, T. K.; Meenakshisundaram, V.; Hung, J.-H.; Simmons, D. S. Neural-network-biased genetic algorithms for materials design: Evolutionary algorithms that learn. *ACS Comb. Sci.* **2017**, *19* (2), 96–107.
- (59) Wertheim, M. Exact solution of the Percus-Yevick integral equation for hard spheres. *Phys. Rev. Lett.* **1963**, *10* (8), 321.
- (60) Thiele, E. Equation of state for hard spheres. *J. Chem. Phys.* **1963**, *39* (2), 474–479.
- (61) Martin, T. B.; Jayaraman, A. Using theory and simulations to calculate effective interactions in polymer nanocomposites with polymer-grafted nanoparticles. *Macromolecules* **2016**, *49* (24), 9684–9692.
- (62) Levi, L.; Raim, V.; Srebnik, S. A brief review of coarse-grained and other computational studies of molecularly imprinted polymers. *J. Mol. Recognit.* **2011**, *24* (6), 883–891.
- (63) Kotelyanskii, M. Simulation methods for modeling amorphous polymers. *Trends Polym. Sci.* **1997**, *5* (6), 192–197.
- (64) Karatrantos, A.; Clarke, N.; Kröger, M. Modeling of polymer structure and conformations in polymer nanocomposites from atomistic to mesoscale: A Review. *Polym. Rev.* **2016**, *56* (3), 385–428.
- (65) Kremer, K.; Müller-Plathe, F. Multiscale simulation in polymer science. *Mol. Simul.* **2002**, *28* (8–9), 729–750.
- (66) Ceperley, D.; Kalos, M. H.; Lebowitz, J. L. Computer Simulation of the Dynamics of a Single Polymer Chain. *Phys. Rev. Lett.* **1978**, *41*, 313–316.
- (67) 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*, 3628–3631.
- (68) 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.
- (69) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (70) Su, L.; Khan, S.; Fan, J. W.; Lin, Y. N.; Wang, H.; Gustafson, T. P.; Zhang, F. W.; Wooley, K. L. Functional sugar-based polymers and nanostructures comprised of degradable poly(D-glucose carbonate)s. *Polym. Chem.* **2017**, *8* (10), 1699–1707.
- (71) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9* (7), 676.
- (72) Kline, S. R. Reduction and analysis of SANS and USANS data using IGOR Pro. *J. Appl. Crystallogr.* **2006**, *39* (6), 895–900.
- (73) Doucet, M.; Cho, J. H.; Alina, G.; Bakker, J.; Bouwman, W.; Butler, P.; Campbell, K.; Gonzales, M.; Heenan, R.; Jackson, A. *SasView Version 4.1* **2017**, DOI: 10.5281/zenodo.
- (74) Heymann, A.; Sinn, C.; Palberg, T. Particle characterization using multiple scattering decorrelation methods: Hard-sphere model

system. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, 62 (1), 813.

(75) Cousin, F. In *Small angle neutron scattering, EPJ. Web of Conferences*; EDP Sciences: 2015; p 01004.

(76) Kamide, K.; Saito, M.; Suzuki, H. Persistence length of cellulose and cellulose derivatives in solution. *Makromol. Chem., Rapid Commun.* **1983**, 4 (1), 33–39.

(77) Khougaz, K.; Gao, Z.; Eisenberg, A. Determination of the critical micelle concentration of block copolymer micelles by static light scattering. *Macromolecules* **1994**, 27 (22), 6341–6346.

(78) Gao, Z.; Eisenberg, A. A model of micellization for block copolymers in solutions. *Macromolecules* **1993**, 26 (26), 7353–7360.

(79) Zhao, C. L.; Winnik, M. A.; Riess, G.; Croucher, M. D. Fluorescence probe techniques used to study micelle formation in water-soluble block copolymers. *Langmuir* **1990**, 6 (2), 514–516.