

# Sequence Patterning, Morphology, and Dispersity in Single-Chain Nanoparticles: Insights from Simulation and Machine Learning

Published as part of the ACS Polymers Au virtual special issue "2023 Rising Stars".

Roshan A. Patel, Sophia Colmenares, and Michael A. Webb\*



Cite This: ACS Polym. Au 2023, 3, 284–294



Read Online

ACCESS |

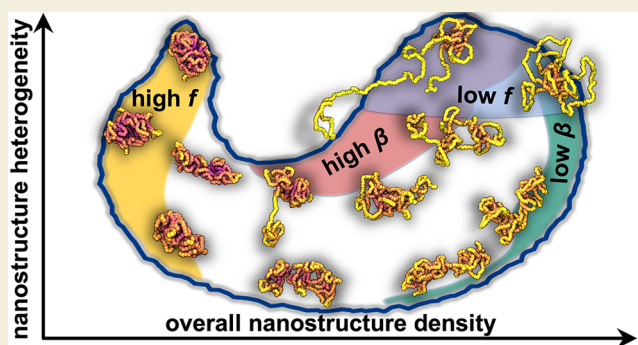
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Single-chain nanoparticles (SCNPs) are intriguing materials inspired by proteins that consist of a single precursor polymer chain that has collapsed into a stable structure. In many prospective applications, such as catalysis, the utility of a single-chain nanoparticle will intricately depend on the formation of a mostly specific structure or morphology. However, it is not generally well understood how to reliably control the morphology of single-chain nanoparticles. To address this knowledge gap, we simulate the formation of 7680 distinct single-chain nanoparticles from precursor chains that span a wide range of, in principle, tunable patterning characteristics of cross-linking moieties. Using a combination of molecular simulation and machine learning analyses, we show how the overall fraction of functionalization and blockiness of cross-linking moieties biases the formation of certain local and global morphological characteristics. Importantly, we illustrate and quantify the dispersity of morphologies that arise due to the stochastic nature of collapse from a well-defined sequence as well as from the ensemble of sequences that correspond to a given specification of precursor parameters. Moreover, we also examine the efficacy of precise sequence control in achieving morphological outcomes in different regimes of precursor parameters. Overall, this work critically assesses how precursor chains might be feasibly tailored to achieve given SCNP morphologies and provides a platform to pursue future sequence-based design.

**KEYWORDS:** topology, unsupervised learning, dimensionality reduction, enzyme mimics, structure control, sequence effects, structure–property relationships



## 1. INTRODUCTION

Single-chain nanoparticles (SCNPs) are an intriguing class of materials obtained by collapsing or folding a polymer chain into a stable nanostructure. The formation and stability of a SCNP is driven by intrachain interactions, which are characteristically noncovalent (e.g., hydrophobic, electrostatic, polar), covalent, or dynamic covalent.<sup>1,2</sup> Inspired by proteins, SCNPs have prospective applications in catalysis, nanomedicine, and biosensing.<sup>2–8</sup> Ultimately, the functionality and utility of a SCNP depends on its morphology, which may determine factors such as hydrophobic character near catalytic active sites.<sup>9</sup> Consequently, fundamental study regarding the structures formed by single polymer chains has implications for the technological advancement of SCNPs<sup>10,11</sup> and understanding of single-chain polymer physics.<sup>12–14</sup>

Significant progress has been made in both developing facile chemical pathways for synthesis of possible precursors (i.e., the initial, unfolded polymer chains) and the characterization of resultant morphologies for SCNPs. Typically, precursors are obtained by first synthesizing a polymer with specific

functional groups and then performing postpolymerization reactions at dilute conditions to promote structure formation.<sup>2</sup> Experimental scattering techniques (e.g., small-angle X-ray and neutron scattering, dynamic light scattering, etc.) have shown that the conformational behavior of SCNPs ranges between a self-avoiding random walk ( $\nu \approx 0.56$ ) and globular ( $\nu \approx 1/3$ ) behavior, depending on the nature of the linking chemistry.<sup>3,12,15–20</sup> Reaction conditions may also be varied or advanced synthetic techniques employed to achieve more precise monomer and functional-group placement and bias formation of target structures in SCNPs.<sup>21–25</sup> Substantial structural dispersity, though, can exist among SCNPs that formed from identical precursors in even polymers with

Received: March 2, 2023

Revised: May 15, 2023

Accepted: May 15, 2023

Published: June 5, 2023



relatively small numbers of chemical moieties.<sup>26</sup> Knowledge relating a precursor to structure formation (and its reliability) in resulting SCNPs will be crucial in achieving target functional properties of these materials.

Molecular dynamics (MD) simulation provides a useful tool to examine the formation of SCNPs under well-controlled conditions and with exquisite structural resolution. Prior MD studies have supported experimental observations while also generating insight regarding both the detailed nanoscopic structure of SCNPs and the mechanisms by which they might form.<sup>11,15,16,20,27–30</sup> For example, good solvent conditions limit contact between distal portions of the polymer chain, thereby promoting intrachain linkages over shorter backbone contour distances and resulting in pearl-necklace-like structures.<sup>16,20,28,29</sup> MD simulations have also highlighted new strategies, such as the use of ring-polymer architectures for the precursor or manipulation of solvent quality<sup>29</sup> to bias formation of more compact, globular structures,<sup>11</sup> which have been experimentally challenging to achieve. Nevertheless, most simulations probe either random or regular sequence patterning of cross-linking moieties on SCNP structure formation.<sup>29</sup> Moreover, although structural dispersity is often noted in simulations of SCNP formation, it is rarely characterized, and its implications in technological applications is largely unexplored. Thus, a comprehensive view of the impact of sequence patterning of cross-linking moieties, a thorough assessment of structural dispersity in SCNP, and examination of their interplay is needed.

Exploring and characterizing the structure–function landscape of polymeric materials is generally nontrivial given the multitude of behaviors enabled by a large chemical and architectural space.<sup>31–35</sup> To contend with this challenge, machine learning (ML) techniques have been increasingly utilized to probe and understand structure–function relationships in soft materials.<sup>36–44</sup> In the context of single polymer chains,<sup>45–47</sup> supervised ML models have been proven effective at relating polymer chain characteristics to average conformational behavior, thereby expediting targeted sequence- and composition-based design tasks. Meanwhile, unsupervised ML algorithms have usefully discriminated among morphological structures formed in many-chain soft materials assembly by noncovalent and supramolecular interactions.<sup>48,49</sup> Collective variables obtained from unsupervised ML can also form the basis for predicting and designing morphology.<sup>50</sup> Overall, these works illustrate the promise of ML to help understand the distribution of morphologies that are accessible by a given precursor and to reveal potential strategies to control that distribution.

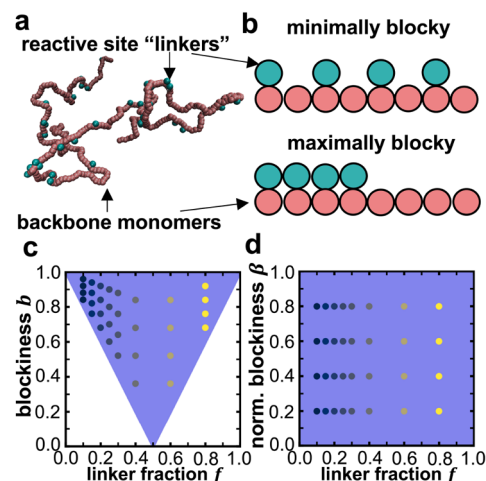
In this study, we use molecular simulation in tandem with machine learning analyses to assess how the patterning of cross-linking moieties on precursor chains impacts the morphology and properties of SCNPs. To comprehensively probe the sequence–structure space, we simulate the formation of 7680 SCNPs from precursor chains that are predominantly distinguished by the fraction and distribution of reactive cross-linking groups along the backbone. These simulations are analyzed to characterize resulting SCNPs via structural analysis and unsupervised manifold learning. These analyses reveal not only the landscape of possible SCNP morphologies but also the dispersity of structures arising from given precursor specifications, which has not been previously well-characterized. We further examine the variability in morphological outcomes for different sequences with the

same patterning descriptors to assess the efficacy of tuning SCNP properties with more precise control over the precursor sequence. In aggregate, this work provides a critical assessment of how precursors might be feasibly tailored to achieve given SCNP morphologies and a platform to pursue SCNP design with specific consideration of structural fidelity.

## 2. METHODS

### 2.1. Systems

**2.1.1. Model of Precursor Chains and Single-Chain Nanoparticles.** We adopt a coarse-grained, phenomenological model to study the formation of SCNPs from precursor polymer chains. Within the model, precursor chains are comprised of backbone beads as well as pendant beads that can function as cross-linking moieties (Figure 1a); these pendant beads are colloquially referenced as “linkers”.



**Figure 1.** Overview of model and precursor chain parameters. (a) Simulation snapshot of a precursor chain. Annotations highlight the presence of backbone beads (pink) and linker beads (teal). Single-chain nanoparticles are formed by allowing the linker beads to react based on specific geometric criteria. (b) Schematic depiction of the precursor parameter blockiness,  $b$ . (c) Visualization of the precursor chain parameter space. Ten distinct precursor chains are generated for each  $(b, f)$  pair indicated by the markers. The underlying shaded region visualizes the accessible space subject to the bound on  $b$  imposed by  $f$ . (d) Visualization of the precursor chain parameter space in the normalized blockiness  $\beta$ – $f$  plane. For a given  $f$ , points are uniformly distributed with respect to  $\beta$ . In panels c and d, the markers are colored according to  $f$  to illustrate the mapping from  $b$  to  $\beta$ .

SCNPs are thus obtained from precursor chains by allowing linkers to react and form bonds. A modified Kremer–Grest model<sup>51</sup> describes the interactions among beads in both precursor chains and SCNPs. The standard elements include (i) excluded-volume interactions characteristic of good-solvent conditions (captured by a Weeks–Chandler–Andersen pair potential) and (ii) covalent-bond interactions (captured by the typical finite extensible nonlinear elastic potential function with standard parameters); good-solvent conditions are often experimentally used to avoid intermolecular linkages between different precursor molecules.<sup>2</sup> The modified components include (i) the introduction of chain semiflexibility, which is common,<sup>52,53</sup> and (ii) the reactivity of linker beads, which has been handled similarly in prior studies of SCNPs.<sup>11,16,20,28</sup> All beads (backbone and linker) are of the same characteristic size  $\sigma$ , the repulsive interaction from the Weeks–Chandler–Andersen pair potential is set by  $\epsilon = k_b T$ , and all bonds are treated equivalently, irrespective of whether they are between backbone–backbone, backbone–linker, or linker–linker beads. Interaction parameters are kept uniform in this study to emphasize the role of patterning of

cross-linking moieties, although manipulation of monomer chemistry and sequence effects to bias structure formation<sup>45</sup> may be of future interest.

Chain semiflexibility is introduced via 1–3 bending interactions of the form

$$E(\theta_{ijk}) = K(1 - \cos \theta_{ijk}) \quad (1)$$

for which  $\theta_{ijk}$  is the angle formed by two consecutive bonds between polymer backbone beads ( $i$  with  $j$  and  $j$  with  $k$ ) and  $K$  controls the local bending rigidity, which we set to  $K = 5k_b T$ . Future phenomenological models may benefit from additional bending and torsional interactions that include the linker beads, but these are not included in the present work.

Reactions between linkers to form SCNPs are handled in the following manner. First, reactions between two linkers results in the formation of a covalent bond; this bond is modeled equivalently as with any other covalent bond in the model. Second, reactions between two linkers can and will proceed only if certain geometric criteria are satisfied. In particular, the distance between two linkers must be within a “capture” radius of  $1.3\sigma$ , and the angle formed by bond vectors adjoining the linker and their respective backbone bead must be  $180 \pm 30^\circ$ . Third, reactions may not proceed between linkers on adjacent backbone beads. Fourth, reactions are irreversible (i.e., covalent bonds are preserved for the duration of the subsequent simulation). Fifth, a linker may only participate in one reaction (i.e., its valency is one and saturates upon formation of a covalent bond with another linker). With respect to prior literature on modeling SCNPs, the utilization of a capture radius is common and follows work from Moreno and collaborators,<sup>16,29</sup> whereas the addition of an angle-based criterion is not ubiquitous but has been employed in atomistic simulations.<sup>27</sup> Here, this constraint and the null reactivity among topologically adjacent linkers is motivated by considering steric and electronic factors of conventional cross-linking chemical moieties.<sup>2</sup> Without this addition, we find that linkers on adjacent backbone beads can undergo facile reaction, which seems generally unphysical given the disposition of cross-linking agents.

**2.1.2. Precursor Chain Parameters.** Many facets of precursor chains (chain length, chemistry, patterning, architecture, etc.) might be feasibly manipulated to control SCNP morphology. Here, only particular parameters of precursor chains are varied while others are common to all precursor chains. Specifically, all precursors consist of  $N_{bb} = 400$  backbone beads arranged in a purely linear topology; each backbone bead may be considered as a monomer or constitutional unit. We choose this number of constitutional units inspired by Lo Verso et al.<sup>29</sup> who noted apparent differences in structures formed by precursors with random sequence patterning versus those formed from regular sequences at this length. In principle, the number of backbone units is another physical parameter that can be tuned and experimentally realized by controlling the extent of polymerization. The primary variable parameters herein are (i) the extent of functionalization or linker fraction  $f = N_l/N_{bb}$  and (ii) a descriptor  $b$ , which relates to the “blockiness” of linker beads as distributed along the polymer chain (Figure 1b). The parameter  $b$  is given by

$$b = \frac{1}{N_{bb}} \left( 1 + \sum_{k=0}^{N_{bb}-2} \mathbf{1}_{k,k+1} \right) \quad (2)$$

where  $k$  and  $k + 1$  are indices of backbone beads along the polymer chain ( $\in [0, N_{bb} - 1]$ ) and  $\mathbf{1}_{i,j}$  is an indicator function, such that

$$\mathbf{1}_{i,j} = \begin{cases} 1, & \text{if } i, j \text{ are both bonded to linker} \\ 1, & \text{if } i, j \text{ are both not bonded to linker} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

In this fashion, a precursor sequence with backbone beads perfectly alternating between having and not having a linker would yield  $b = N_{bb}^{-1}$ , while a sequence with all backbone beads possessing linkers followed by all backbone beads without linkers would yield  $b = 1$ . This blockiness parameter bears some similarity to but is not equivalent to

the correlation parameter used to describe the statistics of random copolymers.<sup>54,55</sup>

An overview of precursor chain parameter combinations examined in this study is shown in Figure 1c. Notably, eq 2 implicitly depends on the linker fraction  $f$ , such that the minimum accessible  $b$  varies with  $f$  (i.e.,  $\min b(f) \approx 12(f - 0.5)$ ). Consequently, precursor chains are studied for  $f \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8\}$ , and values of  $b$  at each composition are chosen such that  $\beta \in \{0.2, 0.4, 0.6, 0.8\}$  where

$$\beta(f) = \frac{b(f) - \min b(f)}{1 - \min b(f)} \quad (4)$$

is a normalized blockiness parameter that can be sampled uniformly irrespective of composition (Figure 1d). We emphasized study at lower  $f$  with the expectation that heterogeneity in SCNP morphologies would diminish at higher  $f$ .

Based on prior work,<sup>40,46</sup> we hypothesized that these parameters would enable study of a vast set of precursor chains that would result in distinct SCNPs. Although a precise connection of  $f$  and  $\beta$  to experimental synthetic conditions may be nontrivial, we expect that such parameters would meaningfully relate to aspects of monomer concentration and reactivity ratios.<sup>56,57</sup> In the long term, advances in sequence-level polymerization may enable more precise definitions.<sup>21–25,58</sup>

**2.1.3. Precursor Chain Sequence Generation.** Ten unique sequences are generated for each combination of  $(f, \beta)$  to assess dispersity of SCNP morphologies as a function of precursor parameters. Thus, in total, 320 unique precursor chains are considered. To generate a given precursor chain sequence at a specified  $(f, \beta)$ , a fixed number of linkers (set by  $f$ ) are first randomly distributed across the polymer backbone, and the resulting initial normalized blockiness  $\beta_i^{(0)}$  is computed. Supposing that  $\beta_i^{(0)} \neq \beta$ , a random pair of backbone beads (one with a linker and one without) is selected, and the linker position is swapped if it would yield  $|\beta_i^{(1)} - \beta| < |\beta_i^{(0)} - \beta|$ . This process is repeated until the trial  $|\beta_i^{(k)} - \beta| < 0.001$ . All sequences generated by this process can be found in the Supporting Information.

**2.1.4. Simulation.** All MD simulations are performed using a modified version of the 3 Mar 2020 distribution of the LAMMPS simulation package.<sup>59</sup> Simulations are performed in reduced units with characteristic quantities of  $m$ ,  $\sigma$ , and  $\epsilon$  for mass, distance, and energy, respectively; the reduced time unit is  $\tau = (m\sigma^2/\epsilon)^{1/2}$ . Simulations correspond to a single polymer chain in implicit solvent. There are no periodic boundary conditions to prevent self-interaction. The simulations therefore correspond to SCNPs forming at infinite dilution. Additional details regarding general simulation protocols, precursor preparation, precursor equilibration, SCNP formation, and SCNP simulation are provided in the Supporting Information.

## 2.2. Analysis

**2.2.1. SCNP Structural Descriptors.** Topological descriptors of the SCNP are assessed via its representation as a molecular graph<sup>60,61</sup>  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ . Here, the beads of the SCNP comprise the set of vertices  $\mathcal{V} = (v_1, \dots, v_{N_{bb}}, v_{N_{bb}+1}, \dots, v_{N_{bb}+N_l})$ , and bonds among beads comprise the set of undirected edges or bonds between beads,  $\mathcal{E}$ ; coordinates of beads are not used for this analysis. Using this framework, we compute two topological descriptors: the number of free backbone beads  $n_f$  and the number of topological domains  $n_d$ ; both of these descriptors have been previously examined in simulation studies to assess SCNP structure.<sup>28,62</sup> For their computation, we adopt the labeling algorithm of Moreno et al. reported in ref 62. In brief, for any pair of bonded linkers  $v_m, v_n$ , the backbone vertices along the directed contour path comprise a set  $\mathcal{D}_{m,n}$ ; if there are  $n_b$  linker–linker bonds, then there will be such sets. Then, a domain  $\mathcal{D}$  is defined as the union of sets, for which every set has nonempty intersection with at least one other set in  $\mathcal{D}$ . Backbone beads that are not found in any path set (i.e., the complement of the set of all backbone beads with the union of all domains) constitute a set of free segments  $\mathcal{F}$ , such that  $n_f = |\mathcal{F}|$  is the cardinality (number of members) of the free-segment set.

SCNP structures are also characterized by shape descriptors derived from their gyration tensor:

$$\mathbf{S} = \frac{1}{N_{\text{bb}}} \sum_{i=1}^{N_{\text{bb}}} (\mathbf{r}_i - \mathbf{r}_{\text{cm}})(\mathbf{r}_i - \mathbf{r}_{\text{cm}})^T \quad (5)$$

where  $\mathbf{r}_i$  is a column vector of the position of the  $i$ th backbone bead,  $\mathbf{r}_{\text{cm}}$  is the center-of-mass position of the SCNP, and  $T$  denotes the transpose. Subsequent diagonalization of eq 5 yields  $\mathbf{S} = \text{diag}(\lambda_1^2, \lambda_2^2, \lambda_3^2)$ , where the diagonal elements are the principal moments of the gyration tensor such that  $\lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2$ . From these, conventional shape descriptors<sup>63</sup> are computed, including the radius of gyration

$$R_g = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2} \quad (6)$$

which indicates the overall size of the SCNP; the acylindricity

$$c = \lambda_2^2 - \lambda_1^2 \quad (7)$$

which measures departure from cylindrical symmetry with respect to the eigenvectors of  $\lambda_2$  and  $\lambda_1$ ; the asphericity

$$b = \lambda_3^2 - \frac{1}{2}(\lambda_1^2 + \lambda_2^2) \quad (8)$$

which measures the overall symmetry of the distribution of beads with respect to three coordinate axes; and the relative shape anisotropy

$$\kappa^2 = \frac{3}{2} \frac{\lambda_1^4 + \lambda_2^4 + \lambda_3^4}{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^2} - \frac{1}{2} \in [0, 1] \quad (9)$$

which assesses departure from spherical symmetric morphologies ( $\kappa^2 = 0$ ). Notably, both  $b$  and  $\kappa^2$  will yield zero for spherically symmetric configurations, while  $b$  will yield zero for any Platonic solid. Ensemble averages of the shape parameters are computed using the final  $10^7$  time steps of the production trajectory.

**2.2.2. Unsupervised Machine Learning.** SCNP morphologies are also identified, distinguished, and compared with the aid of unsupervised ML. The essence of the approach is to discriminate between SCNP morphologies in a data-driven fashion on the basis of collective differences in the local environments of the composite backbone beads; this is largely inspired by methods and analyses by Reinhardt, Statt, and co-workers in the context of colloidal crystals, ice crystals, binary mesophases, and model polymer aggregation.<sup>48,50,64</sup> Here, we consider the local density of backbone beads  $n(r_c) = N/\sigma^3$  to be descriptive of the local environment of a bead, such that a given configuration of a SCNP can be numerically represented by a histogram of such local densities. In this study, local densities are measured only for backbone beads using a spherical cutoff of  $r_c = 6\sigma$ , and histograms possess 40 evenly spaced bins over the range  $[6.63 \times 10^{-3}, 0.239]\sigma^{-3}$ , which spans observed values across all configurations.

The histograms can be constructed for any given configuration of a SCNP to provide a 40-dimensional feature vector  $\mathbf{x}_i$ . To obtain a representative ensemble, 20 distinct configurations for each SCNP (dumped at a frequency of  $2.5 \times 10^6$  time steps) are featurized. This results in 320 sequences  $\times$  24 independently formed SCNPs per sequence  $\times$  20 configurations per SCNP = 153,600 total feature vectors; the 20 configurations taken for each SCNP accounts for structural fluctuations, which may be of interest in certain applications. Subsequently, the uniform manifold approximation and projection (UMAP) algorithm is used to learn a low-dimensional numerical embedding of the SCNP morphologies and the manifold over which they are distributed. To facilitate visualization and interpretation, we simply target a two-dimensional embedding. Hyperparameters related to UMAP include the size of the local neighborhood (set to 200), the overall density of the embedding (minimum distance between points set to 1), and the distance metric for points in the feature space (Euclidean). In effect, this approach generates a mapping  $\mathbb{R}^{40} \rightarrow \mathbb{R}^2$ :  $\text{UMAP}(\mathbf{x}_i) = \mathbf{Z}_i$  where  $\mathbf{Z}_i$  is a coordinate vector in the learned low-dimensional representation of the set of local densities around monomers. As such,  $\mathbf{Z}_i$  provides information on the nanostructural characteristics of SCNPs. We note

that additional descriptions of the local environment, including that used by Reinhardt,<sup>64</sup> as well as those equivalent to atom-centered symmetry functions,<sup>65</sup> were also examined for the embedding described above (see Supporting Information, Figure S2). However, all approaches yielded qualitatively similar organization of morphologies, and local density was thus employed for its ease of interpretability, despite it being the least descriptive of the local environment.

**2.2.3. Metrics of Dispersity and Sequence Variability.** To quantify morphological dispersity in SCNPs and assess the importance of precise sequence control, we introduce a series of so-called dispersity and sequence variability metrics. Dispersity metrics primarily address variability in outcomes of SCNPs for a given precursor chain, while sequence variability metrics address the range of possible outcomes for a given specification of  $f$  and  $\beta$ , which is associated with a set of realizable precursor chains. In the following, a given SCNP is characterized by a tuple  $(c(f, \beta), s_c)$  where  $c$  denotes a precursor chain from the set  $C$  of all possible precursor chains with characteristics  $f$  and  $\beta$ , and  $s_c$  is an element of the set  $\mathcal{S}_c$  of all possible SCNP structures that can form from precursor  $c$ . In addition, the notation  $x \sim Y$  indicates that the variable  $x$  was sampled from the set  $Y$ . In the present study, relevant sets practically correspond to (i) the 24 independent trajectories that yield unique SCNPs from a given precursor chain or (ii) the ten distinct precursor chains generated for each combination of  $f$  and  $\beta$ ; these sets are respectively used in the calculation of expectation values indicated by  $\mathbb{E}_{s_c \sim \mathcal{S}_c}[\cdot]$  and  $\mathbb{E}_{c \sim C}[\cdot]$ .

Dispersity in the size of the SCNPs arising from a given precursor chain is given by

$$\mathcal{D}_{R_g}(c(f, \beta)) \equiv \mathbb{E}_{s_c \sim \mathcal{S}_c}[(\langle R_g \rangle_{s_c} - \mathbb{E}_{s_c \sim \mathcal{S}_c}[\langle R_g \rangle_{s_c}])^2] \quad (10)$$

where  $\langle R_g \rangle_{s_c}$  is the ensemble-average  $R_g$  for a given SCNP; this is the variance in  $\langle R_g \rangle_{s_c}$ , sampled over  $\mathcal{S}_c$ . Dispersity in  $\langle \mathbf{Z} \rangle$  (related to the density of local environments) for SCNPs arising from a given precursor chain is given by

$$\mathcal{D}_{\mathbf{Z}}(c(f, \beta)) \equiv \mathbb{E}[-\log(P(\langle \mathbf{Z} \rangle | c)) | c] \quad (11)$$

where “ $x|c$ ” denotes conditioning the variable  $x$  on  $c$ . This quantity is the entropy of the distribution of  $\langle \mathbf{Z} \rangle$  and measures the differences in the nanostructures of the SCNP that are formed from a given precursor  $c$ .

Because not all manifestations of  $c(f, \beta)$  are expected to yield the same distribution of SCNPs, we define four sequence variability metrics. Sequence variability on the outcome of  $R_g$  at a particular  $f$  and  $\beta$  is defined as

$$\mathcal{S}\mathcal{V}_{R_g}(f, \beta) \equiv \mathbb{E}_{c \sim C}[(\mathbb{E}_{s_c \sim \mathcal{S}_c}[\langle R_g \rangle_{s_c}] - \mathbb{E}_{c \sim C}[\mathbb{E}_{s_c \sim \mathcal{S}_c}[\langle R_g \rangle_{s_c}]])^2] \quad (12)$$

which probes the variance of the mean SCNP sizes formed from  $c$  across all  $c$  in  $C$ . Sequence variability on the dispersity of chain sizes is similarly defined as

$$\mathcal{S}\mathcal{V}_{\mathcal{D}_{R_g}}(f, \beta) \equiv \mathbb{E}_{c \sim C}[(\mathcal{D}_{R_g}(c) - \mathbb{E}_{c \sim C}[\mathcal{D}_{R_g}(c)])^2] \quad (13)$$

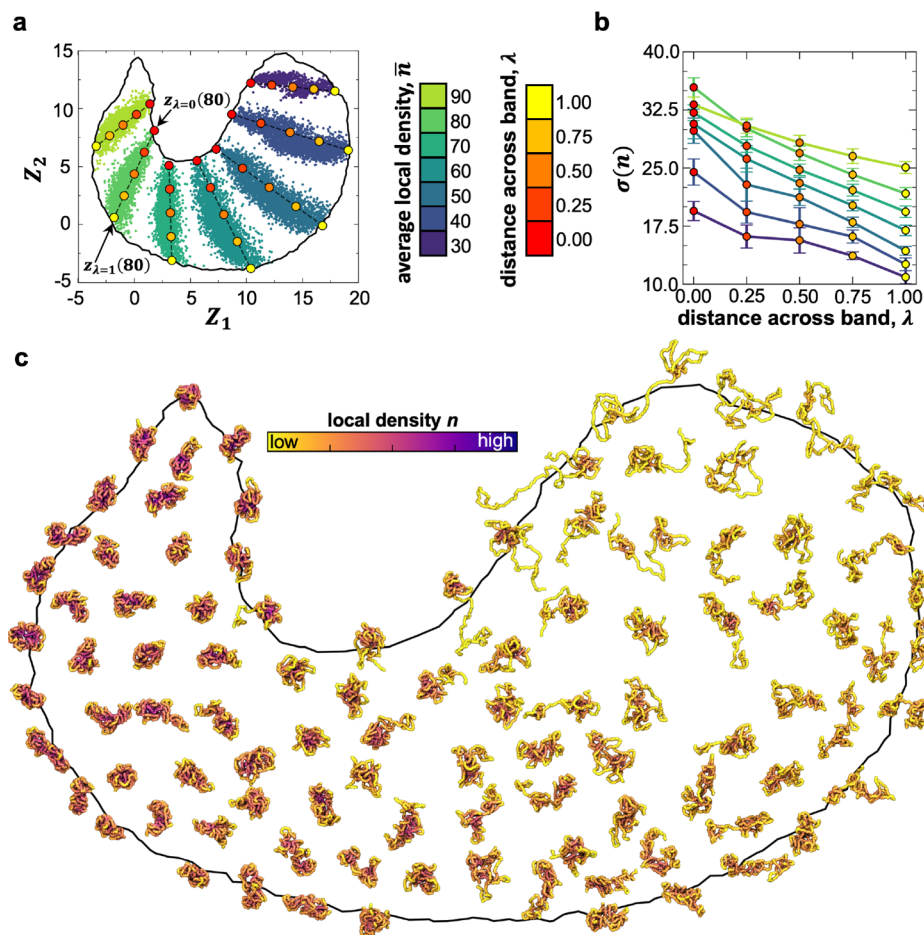
This probes the variance of the dispersities of SCNP sizes (eq 10) formed from  $c$  across all  $c$  in  $C$ . Because the position on the manifold  $\langle \mathbf{Z} \rangle$  is two-dimensional, sequence variability of  $\mathbf{Z}$  is instead defined as

$$\mathcal{S}\mathcal{V}_{\mathbf{Z}}(f, \beta) \equiv \mathbb{E}[-\log(P(\langle \mathbf{Z} \rangle | f, \beta)) | f, \beta], \quad \langle \mathbf{Z} \rangle \equiv \mathbb{E}_{s_c \sim \mathcal{S}_c}[\langle \mathbf{Z} \rangle_{s_c}] \quad (14)$$

This is the entropy of the distribution formed by the expectation values of  $\langle \mathbf{Z} \rangle_{s_c}$  over  $\mathcal{S}_c$  for all  $c$  in  $C$ . Lastly, we define the sequence variability in the dispersity of positions on the manifold:

$$\mathcal{S}\mathcal{V}_{\mathcal{D}_{\mathbf{Z}}}(f, \beta) \equiv \mathbb{E}_{c \sim C}[(\mathcal{D}_{\mathbf{Z}}(c) - \mathbb{E}_{c \sim C}[\mathcal{D}_{\mathbf{Z}}(c)])^2] \quad (15)$$

This is related to the variance in the dispersity of nanostructures (eq 11) formed from  $c$  across all  $c$  in  $C$ .



**Figure 2.** Visualization and analysis of single-chain nanoparticle (SCNP) morphologies. (a) A two-dimensional manifold organization of morphologies for SCNPs based on the Uniform Manifold Approximation and Projection (UMAP) unsupervised learning algorithm. For visual clarity, only a fraction of the overall data is shown (see also Figure S3). Each marker corresponds to a single configuration of a SCNP and is colored by the average local density around backbone beads,  $\bar{n}$ , revealing bands of constant  $\bar{n}$  distributed across the manifold. Five additional markers (warm colors) that correspond to selected coordinates that span the manifold for each band are also shown; these coordinates are used for analysis in panel (b). Particular examples are noted within the figure axes for  $\bar{n} = 80$ . (b) The standard deviation of local densities around backbone beads,  $\sigma(n)$ , as a function of fractional distance  $\lambda$  across each band. The distance across each band is estimated by selection of two distal reference points  $\mathbf{z}_0(\bar{n})$  and  $\mathbf{z}_1(\bar{n})$ , such that intermediate positions can be selected as  $\mathbf{z}_\lambda = (1 - \lambda)\mathbf{z}_0(\bar{n}) + \lambda\mathbf{z}_1(\bar{n})$ . Each marker corresponds to the average of  $\sigma(n)$  for a collection of structures found within 0.3 units of the corresponding marker in panel a; the line colors reflect the band for each  $\bar{n}$ . Error bars report the standard deviation of  $\sigma(n)$  across the aforementioned collection of structures. (c) Simulation snapshots of 103 representative configurations across the latent space manifold. Each rendering is positioned above its coordinate position in the manifold. In the renderings, only backbone beads are shown and are colored by the local density of other backbone beads within a  $6\sigma$  radius.

### 3. RESULTS AND DISCUSSION

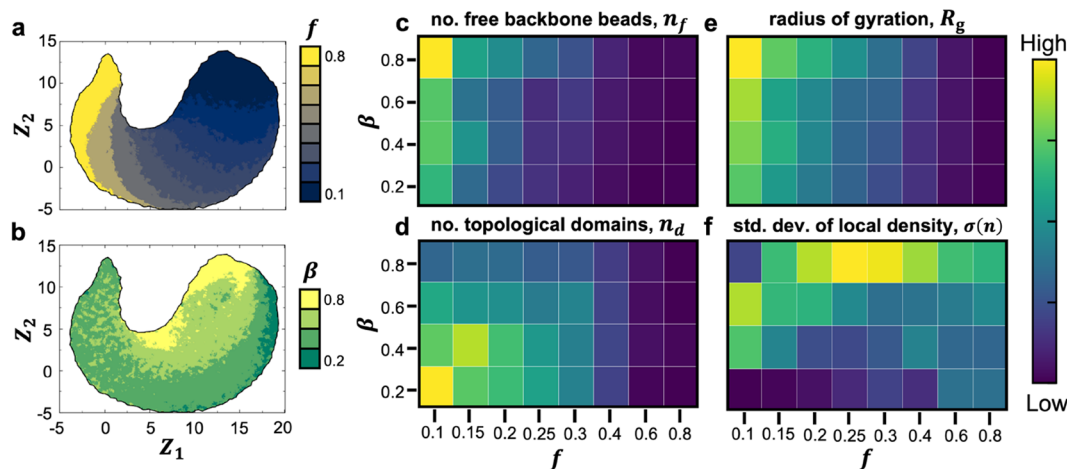
#### 3.1. Characterization of Morphologies

**3.1.1. Unsupervised Learning Facilitates Understanding of SCNP Morphologies.** To effectively process and identify morphological archetypes of all simulated SCNPs, we utilized unsupervised machine learning to construct a low-dimensional manifold of SCNP morphologies (section 2.2) wherein the structure of every SCNP can be represented via a two-dimensional coordinate vector  $\mathbf{Z} = (Z_1, Z_2)$ . Importantly, the morphologies of SCNPs are expected to share similar features when closely positioned on the manifold (e.g., when  $|Z_i - Z_j|$  is small).

Figure 2a shows that the manifold of SCNP morphologies, as determined by the unsupervised machine learning, is primarily organized by the average local density of backbone beads  $\bar{n}$ . Moving from left-to-right (increasing  $Z_1$ ), SCNPs transition from compact (high  $\bar{n}$ , green colors) to expanded

(low  $\bar{n}$ , blue colors) structures. Figure 2b reveals a secondary level of organization, more so along  $Z_2$ , wherein structures are differentiated by the distribution of local environments, which is quantified by the standard deviation of local densities around backbone beads,  $\sigma(n)$ . Moving from the bottom toward the top of the manifold within a band of constant  $\bar{n}$ , morphologies transition from having relatively homogeneous local environments toward having increasingly heterogeneous nanostructures. Thus, the unsupervised learning approach distinguishes between, for example, different manifestations of similarly compact SCNP morphologies (i.e., dense core and less confined shell versus more homogeneous structure).

Figure 2c illustrates how representative SCNP structures are distributed across the low-dimensional manifold, revealing several archetypal morphologies that arise in congruence with the specific local environments. Specifically, structures on the left side of the manifold ( $Z_1 < 5$ ) are globular, typified by relatively large  $\bar{n}$ . Structures on the top-right side of the



**Figure 3.** Relationship between sequence characteristics of precursors and structural features of resulting single-chain nanoparticles. Variation of (a) linker fraction  $f$  and (b) normalized blockiness  $\beta$  as a function of manifold coordinates  $Z_1$  and  $Z_2$ . In both panels a and b, grid points within the manifold are colored by the median value of single-chain nanoparticles within a small radius of the grid point. Heat maps showing the average over single-chain nanoparticles generated from precursors at specified  $\beta$  and  $f$  for (c) the number of free backbone beads  $n_b$ , (d) the number of topological domains  $n_d$ , (e) the radius of gyration  $R_g$ , and (f) the standard deviation of the distribution of local densities around monomer beads. The values used to color (f) are normalized by the mean value observed at a given  $f$  to highlight trends with respect to  $\beta$ .

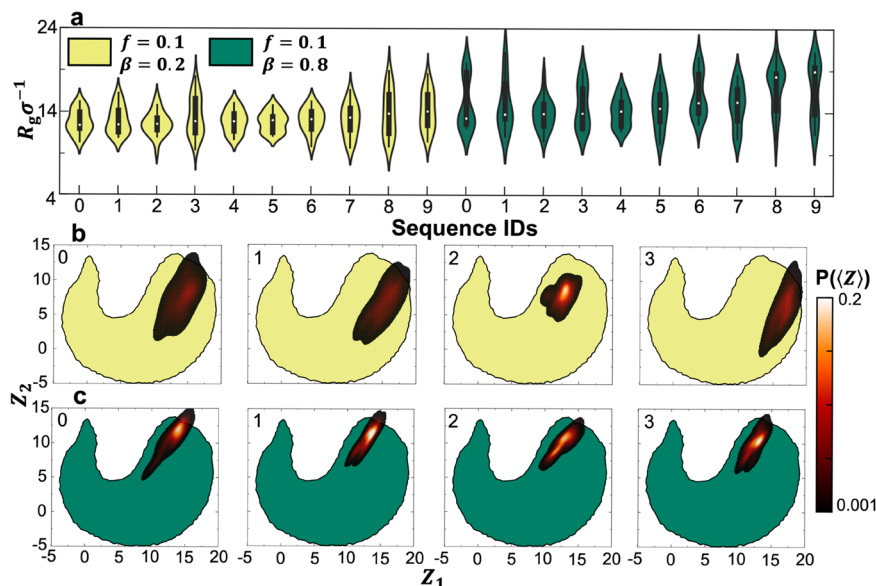
manifold ( $Z_1 \approx 5$ ,  $Z_2 > 5$ ) resemble pearl-necklace and tadpole-like morphologies, typified by larger  $\sigma(n)$ . Structures on the bottom-right side ( $Z_1 > 10$ ,  $Z_2 \approx 0$ ) are populated with diffuse globular, worm-like, and rod-like morphologies, with lesser  $\sigma(n)$ . Therefore, it becomes possible to make a distinction between stringy versus clumpy pearl-necklace-like structures based on the heterogeneity of local environments. Collectively, these results indicate that (i) the precursor chains over the  $\beta$ - $f$  parameter space yield a rich array of SCNP morphologies and (ii) the manifold coordinates  $Z$ , which broadly correlate to the average and heterogeneity of nanostructural environments within a SCNP, provide a reasonable organization of those structures.

Observed archetypes also resemble those of prior studies, albeit with some distinctions. While Moreno and co-workers observe many of the same archetypes, they noted an abundance of pearl-necklace-like structures<sup>16,20,28</sup> which are represented to a much lesser degree across our morphologies. We suggest that this difference arises due to our use of angular constraints for linker-bond formation, which inhibits linkers forming bonds over short contour distances of the polymer backbone. In another study, Tulsi and Simmons found SCNPs that also adopt pearl-necklace, worm-like, and dense globular structures.<sup>30</sup> They further identified “gnarled” structures, which are similar to the more diffuse globules observed in this study. Interestingly, their SCNPs arise from a model with different physical characteristics and distinct mechanism for assembly; this suggests that the present and prior theoretical studies may be homing in on a general set of readily realizable SCNP morphologies. In addition, many of our structures resemble tadpole morphologies which are similar to morphologies observed in experiments.<sup>66–68</sup>

**3.1.2. Sequence Characteristics of Precursors Bias Structure Formation.** To assess how linker patterning impacts SCNP structure, we examine how precursor parameters  $f$  and  $\beta$  dictate position on the manifold of SCNP morphologies (Figure 3a,b). While  $f$  and  $\beta$  both impact morphology, they have different effects. Roughly, trends with  $f$  and  $Z$  track with  $\bar{n}$ . This suggests that structures formed from

high- $f$  precursors (left) predominantly display more compact, globular morphologies typified by high  $\bar{n}$ , while structures formed from low- $f$  precursors (right) result in more diffuse, low- $\bar{n}$  morphologies. The impact of  $\beta$  is less striking but evident at low  $f$  (i.e.,  $f \lesssim 0.3$ ). In particular,  $\beta$  impacts how cross-links are distributed over the SCNP topology: at fixed and low  $f$ , high- $\beta$  precursors yield SCNPs with few centralized “hubs” of cross-links, while the low- $\beta$  precursors result in SCNPs with cross-links that are more evenly distributed over the chain. As a result, increasing  $\beta$  biases morphologies from worm- or rod-like (bottom right) to more pearl-necklace- and tadpole-like structures (top right); the nature of this effect is further resolved in Figure S4.

We characterized 7680 unique SCNPs across the  $\beta$ - $f$  space by the number of free backbone beads  $n_b$ , the number of domains  $n_d$ , and the radius of gyration  $R_g$  to quantitatively examine how particular morphologies arise as a result of  $\beta$  and  $f$ . Figure 3c–f show the averages of these structural descriptors over the ensemble of SCNPs generated from precursors with specific blockiness and linker-fraction combinations; precise numerical values are reported in the Supporting Information. All four descriptors possess identifiable trends that resonate with the qualitative observations from Figures 2c and 3a,b. Figure 3c shows that decreasing  $f$  and increasing  $\beta$  typically increases the extent of polymer chain that is not present in any cross-linked domain. Meanwhile, Figure 3d indicates that decreasing  $\beta$  and  $f$  tends to result in SCNPs with more domains that possess fewer beads. Together, these results suggest that  $\beta$  and  $f$  can be manipulated to tune aspects of SCNP topology, although not arbitrarily. Beyond a certain  $f$ , it appears these structural descriptors are not strongly affected by  $\beta$ . This is evident by the flat color saturation at  $f \geq 0.6$ , indicating that all SCNPs are essentially one large topological domain in this precursor regime. The nature of these topological structures has clear implications on the size of the SCNP, as seen in Figure 3e. To first order, SCNP size, as given by  $R_g$ , is controlled by the number of free segments, which is strongly biased by  $f$ ; only at low  $f$  does  $\beta$  appear to have subtle impact on  $R_g$ ; we observe near-identical trends for



**Figure 4.** Morphological dispersity of single-chain nanoparticles originating from given precursor sequences. (a) Comparison of distribution of radius of gyration  $P(\langle R_g \rangle)$  for distinct single-chain nanoparticles. The distributions are obtained from 24 independent replicate simulations of the same precursor chain sequence. The data are for SCNPs formed from ten precursor sequences each from  $f = 0.1$  and  $\beta = 0.2$  (yellow, left) and  $f = 0.1$  and  $\beta = 0.8$  (green, right). The width of violins correspond to the density obtained from Gaussian kernel density estimation; the edges of box plots in the violin depict the interquartile range, while white dots indicate the median value. Comparison of the distribution of manifold-coordinate vectors  $P(\langle Z \rangle)$  for single-chain nanoparticles formed from selected precursor chain sequences with parameters of (b)  $f = 0.1$  and  $\beta = 0.2$  and (c)  $f = 0.1$  and  $\beta = 0.8$ . In panels b and c, the color reflects Gaussian kernel density estimation over the  $\langle Z \rangle$  for each of the 24 distinct SCNPs formed by each precursor sequence. The color scheme reflecting precursor parameters and sequence labels is the same across panels.

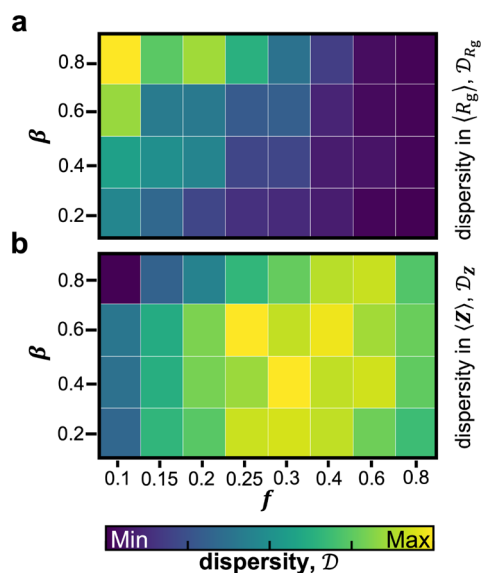
shape descriptors derived from the gyration tensor (Figure S5). Rather,  $\beta$  mostly controls the extent of heterogeneity of SCNP nanostructures (Figure 3f). In particular, decreasing  $\beta$  biases from morphologies that feature more heterogeneous local environments around beads toward those with more homogeneous ones, even while  $R_g$  is largely unaffected. Collectively, these results illustrate how sequence patterning of precursors can be manipulated to bias the morphological characteristics of SCNPs.

### 3.2. Characterization of Structural Dispersity

**3.2.1. Morphological Dispersity Depends on Sequence Patterning.** We next investigate the impact of precursor patterning on the consistency of forming SCNP morphologies. Figure 4 illustrates that a single precursor sequence can give rise to a set of SCNPs with diverse morphological characteristics. This is first demonstrated in Figure 4a, which compares  $P(\langle R_g \rangle)$  for ten SCNPs originating from precursor chains with  $f = 0.1$  and  $\beta = 0.2$  and ten originating from precursor chains with  $f = 0.1$  and  $\beta = 0.8$ . Selected sequences (e.g., 5, 6, and 9 for  $f = 0.1$  and  $\beta = 0.8$ ) are noted to have particularly broad distributions, which result from averaging over the 24 replicate simulations. Across both precursor conditions, many distributions are non-Gaussian, bordering on bimodal or heavy-tailed, which would indicate that the different trajectories produce disparate morphologies rather than exhibit a single, dominant archetypal morphology. Figure 4b,c shows that there can also be substantial nanostructural diversity in the ensemble SCNPs formed by a given precursor sequence; this would be evidenced by a diffuse  $P(\langle Z \rangle)$  for a given sequence. Such diffuse distributions are particularly identifiable for sequence 0, 1, and 3 in Figure 4b, which derive from low-blocky precursors. However, sequence patterning can significantly impact the tightness of  $P(\langle Z \rangle)$ , as

sequence 2 for  $\beta = 0.2$  is much more strongly peaked than that of sequence 0. Diffuse  $P(\langle Z \rangle)$  for  $\beta = 0.8$  (Figure 4c) are not quite as evident, and the  $P(\langle Z \rangle)$  are generally similar in shape, hinting at the possible role of precursor parameters like  $\beta$  on dispersity.

**3.2.2. Precursor Parameters Have Disparate Effects on Different Measures of Dispersity.** Figure 5 elucidates the role of  $f$  and  $\beta$  on both dispersity in  $P(\langle R_g \rangle)$  and dispersity in  $P(\langle Z \rangle)$ . Intriguingly, we find that these measures of dispersity are generally anticorrelated. This is evident by examining trends with increasing  $f$ , which tends to decrease dispersity in  $P(\langle R_g \rangle)$  and increase dispersity in  $P(\langle Z \rangle)$ . This can be understood by considering the high- $f$  regime, for which well-connected globular morphologies would have similar overall size and limited capacity to change (low dispersity in  $P(\langle R_g \rangle)$ ) but can nonetheless have disparate populations of local environments (low dispersity in  $P(\langle Z \rangle)$ ), such as observed in Figure 2. At low  $f$  but high  $\beta$ , dispersity in  $P(\langle R_g \rangle)$  is high yet dispersity in  $P(\langle Z \rangle)$  is diminished. Combined with the observation that SCNPs formed from precursors with these characteristics resemble pearl-necklace-like morphologies, diminished dispersity in  $P(\langle Z \rangle)$  suggests that the formation of dense (pearl) and diffuse (lace) nanostructures is relatively consistent. Furthermore, heightened dispersity in  $P(\langle R_g \rangle)$  suggests that how these nanostructures are arranged or connected (i.e., the precise topology) could result in relatively large changes  $P(\langle R_g \rangle)$ . In addition, we observe an apparent maximum in  $P(\langle Z \rangle)$  dispersity at intermediate  $f$  (0.25, 0.3) and intermediate  $\beta$  (0.4, 0.6), suggesting relatively greater diversity in the set of nanostructures between SCNPs formed from these precursors. Moving away from this maximum toward higher  $f$  decreases the entropy of the distributions but to a lesser degree than moving away from the maximum toward lower  $f$ . This reflects

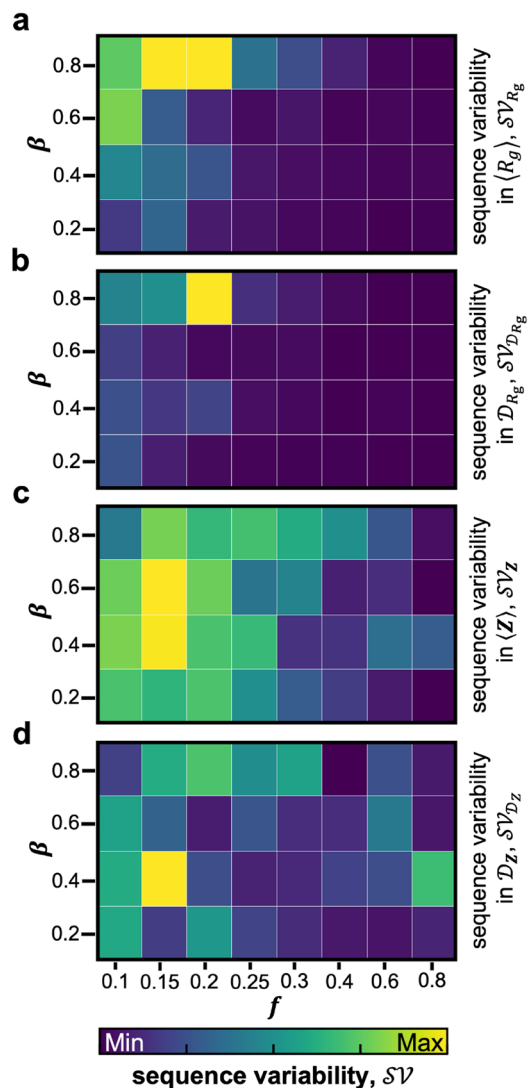


**Figure 5.** Effect of precursor parameters on morphological dispersity. Heat maps at a given linker fraction  $f$  and normalized blockiness  $\beta$  for the average (a) dispersity in  $\langle R_g \rangle$  given by  $\mathcal{D}_{R_g}$  in eq 10 and (b) dispersity in manifold coordinates given by  $\mathcal{D}_Z$  in eq 11. For both panels, average effects are estimated by considering the ten unique precursor chains generated at a given  $f$  and  $\beta$ .

the consistent, high-density core formed by structures at high  $f$  but high variability in local densities around the outside of the structure that depends on the precise folded shape of the SCNP even when broadly characterized as a compact globule.

**3.2.3. Specific Sequences Enable More Precise Control over Morphology.** Differences in the distributions of  $\langle R_g \rangle$  and  $\langle Z \rangle$  across sequences with the same  $\beta$  and  $f$  exhibit differences (Figure 4) suggesting that sequences might be precisely crafted to fine-tune characteristics of formed morphologies. To examine this, we calculate four “sequence variability” metrics that report how distinct aspects of SCNP morphologies are across different sequences for the same  $f$  and  $\beta$ . In particular, we specifically examine the sequence variability in the mean of  $P(\langle R_g \rangle)$ , the dispersity in  $P(\langle R_g \rangle)$ , the mean of  $P(\langle Z \rangle)$ , and the dispersity in  $P(\langle Z \rangle)$ . We anticipate that understanding sequence variability may be useful to ascertain whether tailoring sequences would offer prospective advantages over an ensemble of sequences defined by a given  $f$  and  $\beta$ .

Figure 6a,b shows that sequence variability is highest for the mean value and dispersity of  $P(\langle R_g \rangle)$  at low  $f$  and high  $\beta$  values. By contrast, sequence variability in the mean and dispersity of  $P(\langle R_g \rangle)$  is diminished at high  $f$ , suggesting that precise sequence control would be less valuable; this observation is consistent with prior results (Figure 3 and Figure 5) that indicated negligible impact of  $\beta$  on structure formation at high  $f$ . Furthermore, Figure 6c indicates that sequence variability in the mean of  $P(\langle Z \rangle)$  follows a similar trend, peaking at small  $f$  and intermediate values of  $\beta$  and decreasing at large  $f$ . There are no discernible trends in sequence variability for the dispersity of  $P(\langle Z \rangle)$  (Figure 6d). Together, these results suggest that precisely tailoring sequences to manipulate the distribution of formed SCNPs, whether it be aspects of the SCNP size or its average local environment, is likely to be most effective for low  $f$  and intermediate to high  $\beta$ .



**Figure 6.** Analysis of sequence effects on morphological control. Heat maps show sequence variability as a function of linker fraction  $f$  and normalized blockiness  $\beta$  calculated as (a) variance in the means of  $P(\langle R_g \rangle)$ , (b) variance in dispersities of  $P(\langle R_g \rangle)$ , (c) entropy of the means of  $P(\langle Z \rangle)$ , and (d) variance in the dispersity of  $P(\langle Z \rangle)$ . Quantities are computed from 10 independently generated sequences for each specified  $f$  and  $\beta$ .

## 4. CONCLUSIONS

In this work, we examined how sequence patterning of intramolecular cross-linking moieties (linkers) in polymer chains impacts the formation of single-chain nanoparticles (SCNPs). To do so, we simulated the formation of 7680 unique SCNPs from precursor chains, which altogether comprehensively spanned a parameter space defined by the number of linkers in the chain,  $f$ , and the blockiness of the chain,  $\beta$ . The morphologies of the SCNPs were subsequently characterized using unsupervised machine learning and several structural descriptors to elucidate the general roles of  $\beta$  and  $f$  on morphological outcomes of SCNPs. Finally, we assessed how  $\beta$  and  $f$  impact dispersity in the sizes of SCNPs ( $P(\langle R_g \rangle)$ ) and their characteristic nanostructural environments ( $P(\langle Z \rangle)$ ), and we explored how precise sequence control might be useful to further tune SCNP properties.



Overall, we found several major trends with respect to SCNP topology, morphology, and precursor parameters. Using manifold learning, we showed that SCNP morphologies could be distinguished principally by the mean-density of beads and secondarily by the heterogeneity of local nanostructural environments. For example, we identified two distinct globular morphologies (one with a dense core–flexible shell structure and one with more homogeneous density throughout). Specific archetypal morphologies were also found to be strongly associated with certain regimes of precursor parameters. Low- $f$  precursors tend to adopt topologically diffuse structures with several domains and large chain segments between domains; in this regime, increasing  $\beta$  biases structure formation toward fewer and larger domains but a higher proportion of the polymer chain that is not within any topological domain. By contrast, precursors with high  $f$  consistently give rise to SCNPs with globular structures that show weak sensitivity to  $\beta$ . In addition, while low- $\beta$  precursors tend to result in SCNPs with rod-like morphologies (i.e., several domains of similar nanostructures), high- $\beta$  precursors result in tadpole-like morphologies (i.e., fewer domains with disparate nanostructures). By examining the set of structures formed from every precursor, we found that low- $f$  precursors yield SCNPs with substantial diversity in size but overall consistent local environments, while high- $f$  precursors generate SCNPs with the opposite trend. Furthermore, comparing the distributions of SCNPs generated from specific sequences that all correspond to the same  $\beta$  and  $f$  highlighted the potential to leverage precise sequence patterning for morphological control, manifest in either specific average values or distribution characteristics. Investigation of new “sequence variability” metrics revealed that precursor chains with low  $f$  and large  $\beta$  illustrate the widest range of possible morphology and dispersity outcomes. Therefore, we hypothesize that this regime could benefit most from precise sequence control to tailor the morphological properties of SCNPs.

Ultimately, the methods and analyses herein may spawn several directions of future research. First, while the present and many prior studies emphasized the final SCNP morphologies, future work may aim to quantify the physical implications of SCNP structure and understand their relationship to other SCNP properties (e.g., response to shear flow, mechanical unfolding, etc.). Furthermore, there is significant opportunity to establish how precursor patterning dictates the mechanism(s) or pathway(s) of SCNP formation. Examination of formation pathways may facilitate additional understanding as to how SCNP morphology might be more tightly controlled. Finally, our results point to the possibility of tuning the dispersity of SCNP morphologies in numerous ways. While prior work has leveraged assembly protocols to bias structure formation (e.g., toward globular morphologies<sup>29</sup>), here we show that dispersity outcomes (either SCNP size or nanostructural environments) might be manipulated either by setting particular precursor parameters or by tailoring sequences. Furthermore, with many demonstrations of using statistical ensembles of chains in functional materials,<sup>69–74</sup> it will be interesting to consider how distributions of chains can be designed to achieve target properties. Approaches to navigate such design tasks, particularly in the context of experimentally verifiable systems will be needed.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Derived data associated with single-chain nanoparticles, their characteristics, and accompanying Python scripts are available for download at [https://github.com/webbtheosim/scnp\\_public](https://github.com/webbtheosim/scnp_public). The configurations of all single-chain nanoparticles used for analysis in this manuscript is published in ref 75.

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acspolymersau.3c00007>.

A complete data set including SCNP properties and list of SCNP precursor sequences (ZIP)

Information regarding simulation protocols, additional strategies explored for unsupervised learning, correlation between topological descriptors, and correlation between topological and morphological descriptors. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Michael A. Webb** – Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA; [orcid.org/0000-0002-7420-4474](https://orcid.org/0000-0002-7420-4474); Email: [mawebb@princeton.edu](mailto:mawebb@princeton.edu)

### Authors

**Roshan A. Patel** – Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA

**Sophia Colmenares** – Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acspolymersau.3c00007>

### Author Contributions

CRediT: **Roshan A. Patel** conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), software (lead), supervision (supporting), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (supporting); **Sophia Colmenares** conceptualization (supporting), data curation (supporting), formal analysis (supporting), investigation (supporting), methodology (supporting), software (supporting), writing-original draft (supporting), writing-review & editing (supporting); **Michael A. Webb** conceptualization (lead), data curation (supporting), formal analysis (supporting), funding acquisition (lead), investigation (supporting), methodology (equal), project administration (lead), resources (lead), software (supporting), supervision (lead), validation (equal), visualization (equal), writing-original draft (supporting), writing-review & editing (lead).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

R.A.P. and M.A.W. both acknowledge support from the National Science Foundation. The development of certain machine learning approaches and analyses is based on work supported by the National Science Foundation under Grant No. 2118861. The investigation of controlled structure

formation in polymers is supported by the National Science Foundation under Grant No. 2237470. Simulations and analyses were performed using resources from Princeton Research Computing at Princeton University, which is a consortium led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing.

## REFERENCES

- (1) Mavila, S.; Eivgi, O.; Berkovich, I.; Lemcoff, N. G. Intramolecular Cross-Linking Methodologies for the Synthesis of Polymer Nanoparticles. *Chem. Rev.* **2016**, *116*, 878–961.
- (2) Lyon, C. K.; Prasher, A.; Hanlon, A. M.; Tuten, B. T.; Tooley, C. A.; Frank, P. G.; Berda, E. B. A brief user's guide to single-chain nanoparticles. *Polym. Chem.* **2015**, *6*, 181–197.
- (3) Perez-Baena, I.; Barroso-Bujans, F.; Gasser, U.; Arbe, A.; Moreno, A. J.; Colmenero, J.; Pomposo, J. A. Endowing Single-Chain Polymer Nanoparticles with Enzyme-Mimetic Activity. *ACS Macro Lett.* **2013**, *2*, 775–779.
- (4) Liu, Y.; Pujals, S.; Stals, P. J. M.; Paulöhr, T.; Presolski, S. I.; Meijer, E. W.; Albertazzi, L.; Palmans, A. R. A. Catalytically Active Single-Chain Polymeric Nanoparticles: Exploring Their Functions in Complex Biological Media. *J. Am. Chem. Soc.* **2018**, *140*, 3423–3433.
- (5) Rothfuss, H.; Knöfel, N. D.; Roesky, P. W.; Barner-Kowollik, C. Single-Chain Nanoparticles as Catalytic Nanoreactors. *J. Am. Chem. Soc.* **2018**, *140*, 5875–5881.
- (6) Rubio-Cervilla, J.; González, E.; Pomposo, J. Advances in Single-Chain Nanoparticles for Catalysis Applications. *Nanomaterials* **2017**, *7*, 341.
- (7) Chen, R.; Berda, E. B. 100th Anniversary of Macromolecular Science Viewpoint: Re-examining Single-Chain Nanoparticles. *ACS Macro Lett.* **2020**, *9*, 1836–1843.
- (8) Hanlon, A. M.; Lyon, C. K.; Berda, E. B. What Is Next in Single-Chain Nanoparticles? *Macromolecules* **2016**, *49*, 2–14.
- (9) Huerta, E.; Stals, P. J. M.; Meijer, E. W.; Palmans, A. R. A. Consequences of Folding a Water-Soluble Polymer Around an Organocatalyst. *Angew. Chem., Int. Ed.* **2013**, *52*, 2906–2910.
- (10) Liao, S.; Wei, L.; Abriata, L. A.; Stellacci, F. Control and Characterization of the Compactness of Single-Chain Nanoparticles. *Macromolecules* **2021**, *54*, 11459–11467.
- (11) Formanek, M.; Moreno, A. J. Effects of precursor topology and synthesis under crowding conditions on the structure of single-chain polymer nanoparticles. *Soft Matter* **2017**, *13*, 6430–6438.
- (12) Pomposo, J. A.; Perez-Baena, I.; Lo Verso, F.; Moreno, A. J.; Arbe, A.; Colmenero, J. How Far Are Single-Chain Polymer Nanoparticles in Solution from the Globular State? *ACS Macro Lett.* **2014**, *3*, 767–772.
- (13) Schroeder, C. M.; Babcock, H. P.; Shaqfeh, E. S. G.; Chu, S. Observation of Polymer Conformation Hysteresis in Extensional Flow. *Science* **2003**, *301*, 1515–1519.
- (14) Schroeder, C. M. Single polymer dynamics for molecular rheology. *J. Rheol.* **2018**, *62*, 371–403.
- (15) Arbe, A.; Pomposo, J.; Moreno, A.; LoVerso, F.; González-Burgos, M.; Asenjo-Sanz, I.; Iturrospe, A.; Radulescu, A.; Ivanova, O.; Colmenero, J. Structure and dynamics of single-chain nanoparticles in solution. *Polymer* **2016**, *105*, 532–544.
- (16) Lo Verso, F.; Pomposo, J. A.; Colmenero, J.; Moreno, A. J. Multi-orthogonal folding of single polymer chains into soft nanoparticles. *Soft Matter* **2014**, *10*, 4813–4821.
- (17) Hosono, N.; Gillissen, M. A. J.; Li, Y.; Sheiko, S. S.; Palmans, A. R. A.; Meijer, E. W. Orthogonal Self-Assembly in Folding Block Copolymers. *J. Am. Chem. Soc.* **2013**, *135*, 501–510.
- (18) Sanchez-Sanchez, A.; Akbari, S.; Etxeberria, A.; Arbe, A.; Gasser, U.; Moreno, A. J.; Colmenero, J.; Pomposo, J. A. "Michael" Nanocarriers Mimicking Transient-Binding Disordered Proteins. *ACS Macro Lett.* **2013**, *2*, 491–495.
- (19) Sanchez-Sanchez, A.; Akbari, S.; Moreno, A. J.; Verso, F. L.; Arbe, A.; Colmenero, J.; Pomposo, J. A. Design and Preparation of Single-Chain Nanocarriers Mimicking Disordered Proteins for Combined Delivery of Dermal Bioactive Cargos. *Macromol. Rapid Commun.* **2013**, *34*, 1681–1686.
- (20) Moreno, A. J.; Lo Verso, F.; Sanchez-Sanchez, A.; Arbe, A.; Colmenero, J.; Pomposo, J. A. Advantages of Orthogonal Folding of Single Polymer Chains to Soft Nanoparticles. *Macromolecules* **2013**, *46*, 9748–9759.
- (21) Gody, G.; Maschmeyer, T.; Zetterlund, P. B.; Perrier, S. Rapid and quantitative one-pot synthesis of sequence-controlled polymers by radical polymerization. *Nat. Commun.* **2013**, *4*, 2505.
- (22) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341*, 1238149.
- (23) Anastasaki, A.; Nikolaou, V.; Pappas, G. S.; Zhang, Q.; Wan, C.; Wilson, P.; Davis, T. P.; Whittaker, M. R.; Haddleton, D. M. Photoinduced sequence-control via one pot living radical polymerization of acrylates. *Chem. Sci.* **2014**, *5*, 3536–3542.
- (24) Schmidt, B. V. K. J.; Fechner, N.; Falkenhagen, J.; Lutz, J.-F. Controlled folding of synthetic polymer chains through the formation of positionable covalent bridges. *Nat. Chem.* **2011**, *3*, 234–238.
- (25) Roy, R. K.; Lutz, J.-F. Compartmentalization of Single Polymer Chains by Stepwise Intramolecular Cross-Linking of Sequence-Controlled Macromolecules. *J. Am. Chem. Soc.* **2014**, *136*, 12888–12891.
- (26) Kyoda, K.; Yamamoto, T.; Tezuka, Y. Programmed Polymer Folding with Periodically Positioned Tetrafunctional Telechelic Precursors by Cyclic Ammonium Salt Units as Nodal Points. *J. Am. Chem. Soc.* **2019**, *141*, 7526–7536.
- (27) Liu, J. W.; Mackay, M. E.; Duxbury, P. M. Molecular Dynamics Simulation of Intramolecular Cross-Linking of BCB/Styrene Copolymers. *Macromolecules* **2009**, *42*, 8534–8542.
- (28) Moreno, A. J.; Bacova, P.; Lo Verso, F.; Arbe, A.; Colmenero, J.; Pomposo, J. A. Effect of chain stiffness on the structure of single-chain polymer nanoparticles. *J. Phys.: Condens. Matter* **2018**, *30*, 034001.
- (29) Lo Verso, F.; Pomposo, J. A.; Colmenero, J.; Moreno, A. J. Simulation guided design of globular single-chain nanoparticles by tuning the solvent quality. *Soft Matter* **2015**, *11*, 1369–1375.
- (30) Tulsi, D. K.; Simmons, D. S. Hierarchical Shape-Specified Model Polymer Nanoparticles via Copolymer Sequence Control. *Macromolecules* **2022**, *55*, 1957–1969.
- (31) Binder, K.; Butt, H.-J.; Floudas, G.; Frey, H.; Hsu, H.-P.; Landfester, K.; Kolb, U.; Kühnle, A.; Maskos, M.; Müllen, K.; Paul, W.; Schmidt, M.; Spiess, H. W.; Virnau, P. *From Single Molecules to Nanoscopically Structured Materials*; Springer International Publishing, 2013; pp 115–210.
- (32) Liu, A. J.; Grest, G. S.; Marchetti, M. C.; Grason, G. M.; Robbins, M. O.; Fredrickson, G. H.; Rubinstein, M.; Olvera de la Cruz, M. Opportunities in theoretical and computational polymeric materials and soft matter. *Soft Matter* **2015**, *11*, 2326–2332.
- (33) de Pablo, J. J.; et al. New frontiers for the materials genome initiative. *npj Computational Materials* **2019**, *5*, 41.
- (34) Chen, L.; Pilia, G.; Batra, R.; Huan, T. D.; Kim, C.; Kuenneth, C.; Ramprasad, R. Polymer informatics: Current status and critical next steps. *Materials Science and Engineering: R: Reports* **2021**, *144*, 100595.
- (35) Gormley, A. J.; Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nature Reviews Materials* **2021**, *6*, 642–644.
- (36) Ferguson, A. L. Machine learning and data science in soft materials engineering. *J. Phys.: Condens. Matter* **2018**, *30*, 043002.
- (37) Jackson, N. E.; Webb, M. A.; de Pablo, J. J. Recent advances in machine learning towards multiscale soft materials design. *Current Opinion in Chemical Engineering* **2019**, *23*, 106–114.
- (38) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polymers Au* **2022**, *2*, 8–26.
- (39) Bhattacharya, D.; Patra, T. K. dPOLY: Deep Learning of Polymer Phases and Phase Transition. *Macromolecules* **2021**, *54*, 3065–3074.
- (40) Patel, R. A.; Webb, M. A. Data-Driven Design of Polymer-Based Biomaterials: High-throughput Simulation, Experimentation,

and Machine Learning. *ACS Applied Bio Materials* **2023**, DOI: 10.1021/acsbm.2c00962.

(41) Bale, A. A.; Gautham, S. M. B.; Patra, T. K. Sequence-defined Pareto frontier of a copolymer structure. *J. Polym. Sci.* **2022**, *60*, 2100.

(42) Shi, J.; Quevillon, M. J.; Amorim Valença, P. H.; Whitmer, J. K. Predicting Adhesive Free Energies of Polymer–Surface Interactions with Machine Learning. *ACS Appl. Mater. Interfaces* **2022**, *14*, 37161–37169.

(43) Xu, P.; Chen, H.; Li, M.; Lu, W. New Opportunity: Machine Learning for Polymer Materials Design and Discovery. *Advanced Theory and Simulations* **2022**, *5*, 2100565.

(44) Meyer, T. A.; Ramirez, C.; Tamasi, M. J.; Gormley, A. J. A User's Guide to Machine Learning for Polymeric Biomaterials. *ACS Polymers Au* **2023**, *3*, 141–157.

(45) Webb, M. A.; Jackson, N. E.; Gil, P. S.; de Pablo, J. J. Targeted sequence design within the coarse-grained polymer genome. *Science Advances* **2020**, *6*, No. eabc6216.

(46) Patel, R. A.; Borca, C. H.; Webb, M. A. Featurization strategies for polymer sequence or composition design by machine learning. *Molecular Systems Design & Engineering* **2022**, *7*, 661–676.

(47) Upadhya, R.; Tamasi, M.; Mare, E. D.; Murthy, S.; Gormley, A. Data-Driven Design of Protein-Like Single-Chain Polymer Nanoparticles. *ChemRxiv* **2022**.

(48) Statt, A.; Kleeblatt, D. C.; Reinhart, W. F. Unsupervised learning of sequence-specific aggregation behavior for a model copolymer. *Soft Matter* **2021**, *17*, 7697–7707.

(49) Gardin, A.; Perego, C.; Doni, G.; Pavan, G. M. Classifying soft self-assembled materials via unsupervised machine learning of defects. *Communications Chemistry* **2022**, *5*, 82.

(50) Bhattacharya, D.; Kleeblatt, D. C.; Statt, A.; Reinhart, W. F. Predicting aggregate morphology of sequence-defined macromolecules with recurrent neural networks. *Soft Matter* **2022**, *18*, 5037–5051.

(51) Grest, G. S.; Kremer, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A* **1986**, *33*, 3628–3631.

(52) Everaers, R.; Karimi-Varzaneh, H. A.; Fleck, F.; Hojdis, N.; Svaneborg, C. Kremer–Grest Models for Commodity Polymer Melts: Linking Theory, Experiment, and Simulation at the Kuhn Scale. *Macromolecules* **2020**, *53*, 1901–1916.

(53) Nikoubashman, A.; Howard, M. P. Equilibrium Dynamics and Shear Rheology of Semiflexible Polymers in Solution. *Macromolecules* **2017**, *50*, 8279–8289.

(54) Fredrickson, G. H.; Milner, S. T. Thermodynamics of random copolymer melts. *Phys. Rev. Lett.* **1991**, *67*, 835–838.

(55) Fredrickson, G. H.; Milner, S. T.; Leibler, L. Multicritical phenomena and microphase ordering in random block copolymers melts. *Macromolecules* **1992**, *25*, 6341–6354.

(56) Jayapurna, I.; Ruan, Z.; Eres, M.; Jalagam, P.; Jenkins, S.; Xu, T. Sequence Design of Random Heteropolymers as Protein Mimics. *Biomacromolecules* **2023**, *24*, 652–660.

(57) Smith, A. A. A.; Hall, A.; Wu, V.; Xu, T. Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett.* **2019**, *8*, 36–40.

(58) Deacy, A. C.; Gregory, G. L.; Sulley, G. S.; Chen, T. T. D.; Williams, C. K. Sequence Control from Mixtures: Switchable Polymerization Catalysis and Future Materials Applications. *J. Am. Chem. Soc.* **2021**, *143*, 10021–10040.

(59) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(60) Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Review* **2003**, *45*, 167–256.

(61) Trinajstić, N. *Chemical Graph Theory*; Routledge, 2018; p 16.

(62) Moreno, A. J.; Lo Verso, F.; Arbe, A.; Pomposo, J. A.; Colmenero, J. Concentrated Solutions of Single-Chain Nanoparticles:

A Simple Model for Intrinsically Disordered Proteins under Crowding Conditions. *J. Phys. Chem. Lett.* **2016**, *7*, 838–844.

(63) Theodorou, D. N.; Suter, U. W. Shape of unperturbed linear polymers: polypropylene. *Macromolecules* **1985**, *18*, 1206–1214.

(64) Reinhart, W. F. Unsupervised learning of atomic environments from simple features. *Comput. Mater. Sci.* **2021**, *196*, 110511.

(65) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121*, 9759–9815.

(66) Wen, J.; Yuan, L.; Yang, Y.; Liu, L.; Zhao, H. Self-Assembly of Monotethered Single-Chain Nanoparticle Shape Amphiphiles. *ACS Macro Lett.* **2013**, *2*, 100–106.

(67) Wen, W.; Huang, T.; Guan, S.; Zhao, Y.; Chen, A. Self-Assembly of Single Chain Janus Nanoparticles with Tunable Liquid Crystalline Properties from Stilbene-Containing Block Copolymers. *Macromolecules* **2019**, *52*, 2956–2964.

(68) Nitti, A.; Carfora, R.; Assanelli, G.; Notari, M.; Pasini, D. Single-Chain Polymer Nanoparticles for Addressing Morphologies and Functions at the Nanoscale: A Review. *ACS Applied Nano Materials* **2022**, *5*, 13985–13997.

(69) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34*, 2201809.

(70) Kosuri, S.; Borca, C. H.; Mugnier, H.; Tamasi, M.; Patel, R. A.; Perez, I.; Kumar, S.; Finkel, Z.; Schloss, R.; Cai, L.; Yarmush, M. L.; Webb, M. A.; Gormley, A. J. Machine-Assisted Discovery of Chondroitinase ABC Complexes toward Sustained Neural Regeneration. *Adv. Healthcare Mater.* **2022**, *11*, 2102101.

(71) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmüller, E.; Olvera de la Cruz, M.; Xu, T. Random heteropolymers preserve protein function in foreign environments. *Science* **2018**, *359*, 1239–1243.

(72) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33*, 2105707.

(73) Ruan, Z.; Li, S.; Grigoropoulos, A.; Amiri, H.; Hilburg, S. L.; Chen, H.; Jayapurna, I.; Jiang, T.; Gu, Z.; Alexander-Katz, A.; Bustamante, C.; Huang, H.; Xu, T. Population-based heteropolymer design to mimic protein mixtures. *Nature* **2023**, *615*, 251–258.

(74) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; de la Cruz, M. O.; Rolandi, M.; Xu, T. Single-chain heteropolymers transport protons selectively and rapidly. *Nature* **2020**, *577*, 216–220.

(75) Webb, M. A.; Patel, R. A.; Colmenares, S. Data for Configurations and Characteristics of Simulated Single-Chain Nanoparticles, Princeton University, 2023. DOI: 10.34770/a2db-gy35