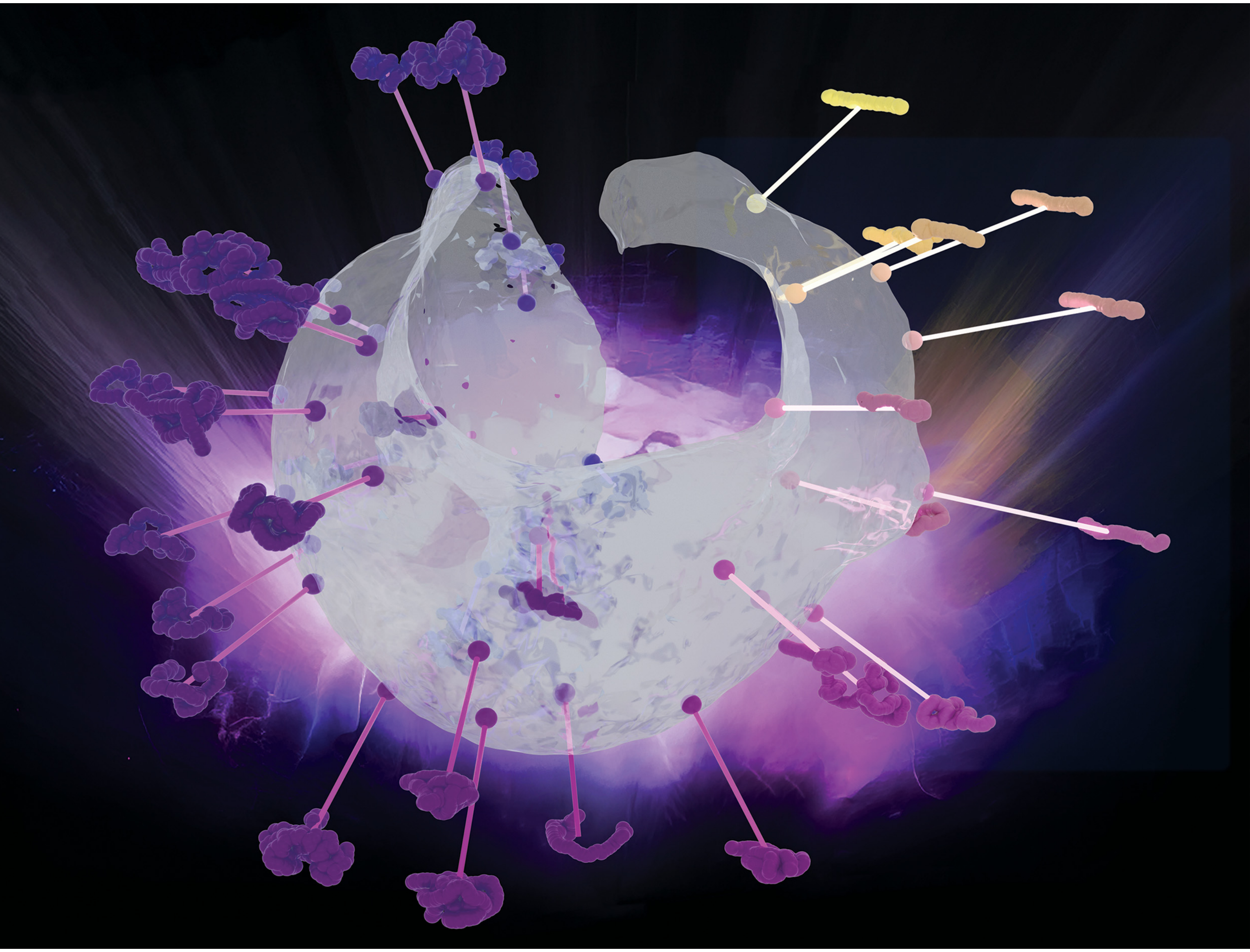
Volume 21
Number 43
21 November 2025
Pages 8245–8390

# Soft Matter

rsc.li/soft-matter-journal



ISSN 1744-6848

ROYAL SOCIETY
OF **CHEMISTRY**

**PAPER**
Howard A. Stone, Michael A. Webb *et al.*
Effect of shear flow and precursor polymer design on single-chain nanoparticle formation

20 YEARS
ANNIVERSARY

Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2025, **21**, 8265

# Effect of shear flow and precursor polymer design on single-chain nanoparticle formation

Matthew D. Chertok,[a] Howard A. Stone *[b] and Michael A. Webb *[a]

Received 17th July 2025,
Accepted 4th October 2025

DOI: 10.1039/d5sm00729a

rsc.li/soft-matter-journal

Single-chain nanoparticles (SCNPs) are a class of materials formed by the intramolecular cross-linking and collapse of single polymer chains. Because their morphology dictates suitability for specific applications, such as nanoscale reactors and drug delivery vehicles, understanding how to control or tailor morphologies is of interest. Here, we investigate how the morphology of SCNPs depends on both precursor chain attributes, such as linker fraction and backbone stiffness, and an imposed shear flow. Using coarse-grained molecular dynamics simulations, we generate an ensemble of structures from 10 800 unique SCNPs, some formed under quiescent conditions and some in shear flow-the latter of which has not been studied previously. We then characterize morphologies by analysis of a three-dimensional embedding space obtained through unsupervised learning of the simulated structures. This reveals how SCNP morphology depends on dimensionless parameters, related to precursor-chain attributes and shear rate, and offers insight into their relative influence. We find that shear rate has comparable influence to the degree of polymerization and the blockiness of reactive sites. Furthermore, shear, which can be externally controlled independent of precursor chain synthesis, can have persistent effects on morphology, such as enhancing compaction of SCNPs based on chain stiffness. This work provides guidelines for designing SCNPs with targeted characteristics based on five dimensionless variables and illustrates the utility of machine learning in analyzing SCNPs formed across a range of conditions.

## 1 Introduction

Single-chain nanoparticles (SCNPs) are polymeric materials formed by intramolecular folding and cross-linking of individual chains, mimicking the compactness, and sometimes function, of natural proteins.[1–4] In one manifestation, precursor chains comprise a polymer backbone bearing side groups (linkers) that can interact *via* non-covalent, covalent, and dynamic-covalent chemistries,[4–6] yielding compartmentalized nanostructures. The resulting materials have numerous prospective applications, including in catalysis,[7–10] $CO_2$ reduction,[11] drug delivery,[12–14] medical imaging,[15] and bio-sensing.[6,9,16–20] A general expectation is that different morphologies of SCNPs may confer distinct functional advantages.

Many morphological outcomes have been highlighted in prior work, ranging from tadpole-like to globular structures.[21] Experimentally, small-angle X-ray and neutron scattering studies have shown that the Flory exponent $\nu$, which characterizes the scaling of the radius of gyration ($R_g$) with backbone length ($N$) as $R_g \propto N^{\nu}$, ranges from $\nu \approx 0.56$ in the self-avoiding coil limit to $\nu \approx 1/3$ in the compact globular limit.[16,22] For example, Vo *et al.* experimentally examined the effect of charge positioning on SCNPs for drug delivery and reported that tadpole-like morphologies produced charge distributions that achieved higher cancer-cell uptake than those exhibited by globular or elongated forms.[23] To effectively design towards specific functions, it is important to understand how tunable precursor parameters (*e.g.*, linker fraction, patterning, chain stiffness, degree of polymerization) influence morphology.

Unlike proteins, whose amino acid sequence deterministically specifies their folded structure,[1] it is important to recognize that SCNPs form by stochastic collapse, yielding a distribution of outcomes, and precursor chain attributes may be likewise characterized by an ensemble of specific chains.[4,24,25] Nevertheless, ensemble-averaged morphologies can be predicted for a given set of precursor attributes, and numerous molecular dynamics (MD) simulations have examined how these attributes bias SCNP structure.[1–4,26–31] These studies describe several notable trends. A higher overall linker fraction promotes uniform backbone compaction, whereas at low linker fractions, blocky arrangements localize cross-linking and produce anisotropic tadpole-like structures with a compact head and an elongated tail.[4] Introducing orthogonal linker species, which react exclusively

---

[a] *Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA. E-mail: mawebb@princeton.edu*
[b] *Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA. E-mail: hastone@princeton.edu*

with a single type, encourages long-range loop formation and yields more spherical particles.[3] Extending the chain length broadens the distribution of $R_g$ without altering the scaling exponent $\nu$.[28] Increased backbone stiffness suppresses short-range loops in favor of larger topological domains, resulting in smaller, more isotropic particles.[27,32] Crowded or poor-solvent conditions drive compaction, particularly in cyclic precursors.[2,26,29] Collectively, these insights establish an initial set of design principles for tuning SCNP morphology under quiescent conditions.

There are several reasons to consider SCNP formation behavior under non-equilibrium conditions. In industrial and biological settings, from extrusion[33] to blood circulation[34] and intracellular biophysical processes,[35] polymer chains are subject to flow, necessitating an understanding of how non-equilibrium conditions affect nanoparticle morphology. Additionally, imposing flow may offer an additional means of morphological control. The conditions imposed by shear are often characterized by the Weissenberg number ($\mathrm{Wi} = \dot{\gamma}\tau_{\mathrm{p}}$), which is a dimensionless parameter that compares the timescale for shear ($\dot{\gamma}^{-1}$) to an intrinsic relaxation time of the polymer ($\tau_{\mathrm{p}}$). Physically, $\mathrm{Wi} \gg 1$ indicates that the chain is deformed more rapidly than it can relax, promoting elongation, while $\mathrm{Wi} \ll 1$ implies that the chain has sufficient time to return toward its equilibrium coil. Using multi-particle collision dynamics (MPCD), Formanek and Moreno investigated the effect of shear flow on fully cross-linked SCNPs.[2,31] For isolated SCNPs, they found that cross-linking constrains extensibility, leading to only modest increases in the radius of gyration with shear rate, scaling as $R_g \sim \mathrm{Wi}^{0.59}$.[31] The dominant structural response to shear was anisotropic deformation with elongation along the flow direction and minor compression in the gradient and vorticity directions. Moreover, the mode of motion under elevated Wi was found to depend on cross-linking density. In particular, sparsely cross-linked SCNPs exhibited tumbling dynamics, while more compact globular forms showed a hybrid behavior combining tumbling with tank-treading, in which the overall shape remains aligned with the flow while monomers circulate around the center of mass. Notably, both studies focused on the response of preformed SCNPs to shear. The question of how shear applied during cross-linking might influence SCNP morphology has been less explored.

Machine learning (ML) provides a powerful framework that may be well-suited for characterizing SCNPs. Broadly, two major categories of ML include supervised learning, where models train on labeled inputs to predict specific targets, and unsupervised learning, where algorithms infer similarity from unlabeled data, automatically grouping samples to uncover emergent patterns.[36] Supervised ML methods have been employed previously to link design parameters of patterned polymers to morphological outcomes. For instance, Webb *et al.* trained deep neural networks to predict $R_g$ of coarse-grained polymer structures, then coupled these models with optimization routines to design new sequences with specific $R_g$.[37] Likewise, Bhattacharya *et al.* showed that recurrent neural networks can predict polymer aggregate morphologies directly from monomer sequences, enabling the design of sequences with targeted aggregation.[38] Such approaches perform well when a clear target metric can be specified. For unsupervised ML, Statt *et al.* applied clustering to MD trajectories of polymer aggregates, revealing a continuous spectrum of assembly structures that conventional order parameters could not resolve.[39] Also, Gardin *et al.* combined rotation- and permutation-invariant fingerprints of monomer environments with density-based clustering to construct a "defectometer" that quantified the formation, healing, and exchange of disordered domains in supramolecular fibers, micelles, and lipid bilayers.[40] In the context of SCNPs, Patel *et al.* used unsupervised ML to group morphologies by similarity of local density histograms, producing a data-driven map of how linker fraction and patterning governed morphology.[4] Collectively, these studies highlight the potential of ML for predicting polymer morphology, albeit without application to non-equilibrium assembly.

In this study, we aim to understand how shear flow, in conjunction with other design parameters, influences the formation and morphology of SCNPs. To explore this, we investigate a five-dimensional parameter space comprising the linker fraction $f$, the chain stiffness parameter $\kappa$ (in units of the basic energy $\varepsilon$), the degree of polymerization $N$, the blockiness $\beta$ of the linker pattern, which quantifies the clustering of reactive beads, and the applied shear rate $\dot{\gamma}$ (Fig. 1). Together, these parameters define a tractable design space for assessing how structural and processing conditions influence SCNP assembly that is also, in principle, experimentally accessible.[25,41–45] We

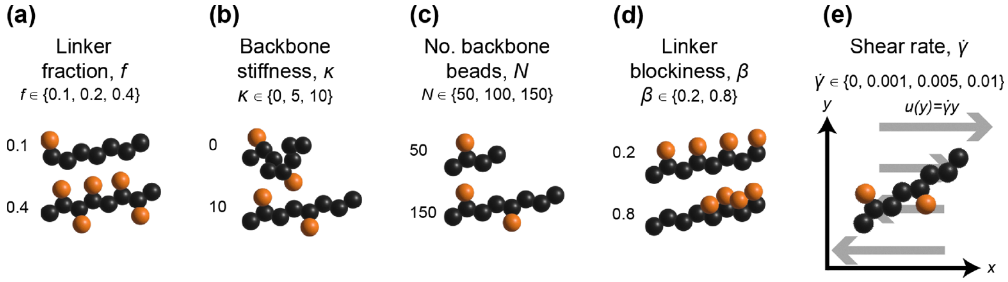


Fig. 1 Overview of precursor chain parameters studied. (a) Linker fraction $f$ is the ratio of reactive linker beads to backbone beads. (b) Stiffness $\kappa$ indicates the tendency for the chain to resist bending. (c) Number of backbone beads $N$. (d) Blockiness $\beta$ quantifies reactive bead clustering. (e) Shear rate $\dot{\gamma}$. Values indicated in each panel are the lower and upper limits of the parameter range in reduced units.

use MPCD to perform 10 800 MD simulations using a phenomenological polymer model[4] across this parameter space, and we organize the dataset with unsupervised ML. This enables analysis on the relative influence of each parameter on morphological outcomes and the interrelationships amongst these variables. Ultimately, this reveals consistent impacts of shear that may also offer practical guidance for future SCNP design.

## 2 Methods

### 2.1 Simulation details

**2.1.1 Phenomenological model of SCNP formation.** All systems are treated at a phenomenological level aimed at capturing essential physics of SCNPs and their formation but without explicit connection to specific chemical manifestations. Precursor polymer chains consist of coarse-grained beads of two types: backbone beads and linker beads. The degree of polymerization is equal to the number of backbone beads, denoted $N$, and the number of linker beads is denoted $N_\ell$. Interactions are generally described using a modified Kremer–Grest framework.[46] In particular, all beads interact *via* the Weeks–Chandler–Andersen potential

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right], & r \le 2^{1/6}\sigma, \\ 0, & r > 2^{1/6}\sigma, \end{cases} \tag{1}$$

where $r$ is the center-to-center distance between two beads. Irreversible reactions between linker beads, which occur based on certain geometric criteria (see simulation procedure), are used to model intramolecular cross-linking of precursor chains to yield SCNPs; following any reaction, the two linker beads are considered directly bonded. Directly bonded beads have a stretching energy given by the finitely extensible nonlinear elastic (FENE) potential

$$U_{\mathrm{bond}}(r) = -\frac{\varepsilon}{2}k_{\mathrm{bond}}\left(\frac{r_0}{\sigma}\right)^2 \ln\left(1 - \left(\frac{r}{r_0}\right)^2\right), \tag{2}$$

where $k_{\mathrm{bond}}$ and $r_0$ are constants. In addition, groups of polymer beads connected by two bonds have a bending energy given by

$$U_{\mathrm{angle}}(\theta) = \kappa\varepsilon(1 + \cos\theta), \tag{3}$$

where $\theta$ is the angle formed by the three interacting polymer beads and $\kappa$ is a parameter that controls polymer stiffness (Fig. 1b); this quantity can also be related to the Kuhn length of the precursor polymer chain (SI, Fig. S1).

Throughout this work, $\sigma$ and $\varepsilon$ are used as characteristic units of length and energy, respectively. Each polymer bead has a characteristic unit mass, $m$. These quantities set the characteristic unit of time to be $\tau = (m\sigma^2/\varepsilon)^{1/2}$. All systems employ $k_{\mathrm{bond}} = 30$ and $r_0 = \frac{3}{2}\sigma$. For polymer stiffness, $\kappa$ is varied across systems and applies to all angles except those between two backbone beads and a linker bead, for which $\kappa/\varepsilon = 0$.

The formation, structure, and dynamics of SCNPs are modeled using the multi-particle collision dynamics (MPCD) formalism.[47,48] Simulations are performed with and without shear flow. MPCD efficiently captures hydrodynamic interactions and thermal fluctuations in a system by using a solvent of momentum-conserving streaming particles. However, these streaming solvent particles do not interact with the polymer beads other than through momentum-exchange. In addition, standard MPCD algorithms do not strictly conserve angular momentum.[47] However, Götze *et al.* showed the lack of angular-momentum conservation negligibly impacts velocity fields in non-rotating, velocity-driven flows (*e.g.*, flow between parallel plates).[49] Consequently, this issue is not expected to have major influence on the results of the present study.

**2.1.2 General simulation protocols.** All simulations are performed using HOOMD-blue version 4.9.[50] Cubic simulation cells are used with edge lengths set as max(100, $N$), where $N$ is the number of backbone beads; this choice mitigates flow artifacts[51] and polymer self-interaction at a reasonable computational expense. Dynamics are integrated with a time step of $\delta t = 0.005\tau$, following Wani *et al.*[52] MPCD simulations use a collision interval of $\tau_{\mathrm{collide}} = 20\delta t$, a collision angle of $\alpha = 130°$, and a cell number density of $\rho = 5/\sigma^3$. The temperature is set such that $k_{\mathrm{B}}T = \varepsilon$, and is regulated using a Maxwell–Boltzmann thermostat.[48,50] Periodic boundary conditions are applied in the $x$ and $z$ directions, while no-slip, reflecting boundaries are imposed in the $y$ direction. Where appropriate, shear flow is induced by translating these boundaries at equal and opposite velocities along the $x$ axis.

Systems are initialized with a single, linear precursor polymer chain at the center of the simulation cell; its end-to-end vector is oriented in the (1, 0, 1) direction. From this initial configuration, simulations are performed for $10^6$ time steps without any cross-linking. During this initial period, configurations are saved every $10^5$ steps to yield ten independent starting configurations per precursor chain. From each of these configurations, simulations are then performed for $2 \times 10^7$ time steps, which is sufficient to allow for near-complete cross-linking and sampling of fully reacted configurations (Fig. S3). During this SCNP-formation period, linker beads are allowed to react. Generally, $10^7$ steps sufficiently determines the morphology of the SCNP, and few reactive events occur after that point (see SI, Fig. S3), which is consistent with the work of Liu *et al.*[30] Thus, in the last $10^7$ time steps of the SCNP-formation period, system configurations are saved every $10^6$ steps for analysis. In total, we consider 216 unique parameter sets ($f$, $\kappa$, $N$, $\beta$, $\dot{\gamma}$), five distinct chains corresponding to each parameter set, and ten replicates per chain. This results in 10 800 simulations and a total of 108 000 SCNP configurations.

**2.1.3 Cross-linking protocol.** During the SCNP-formation period, all pairs of free (*i.e.*, not previously reacted) linkers are evaluated for potential bond formation every 100 integration steps. This interval was selected as a compromise: more frequent evaluations would substantially increase computational cost due to the additional overhead, whereas less frequent evaluations risk missing viable interactions. Bonds are formed based on three criteria.[4,28,29] First, their separation distance must be $r \le 1.3\sigma$. Second, both backbone-linker-linker angles must satisfy $\theta \geq 150^\circ$; this ensures that bonds are only formed when the moieties are at particular relative orientations. Third, each linker is the nearest linker to its prospective eligible partner. If these criteria are satisfied, then the two linker beads are bonded using eqn (2). Fig. S10 illustrates the impact of changing the capture radius and omitting the angle criterion, demonstrating that including the angle criterion promotes globular morphologies by suppressing trivial local reactions in favor of longer-range cross-links.

**2.1.4 Post-crosslinking simulations under zero shear.** In addition to the simulations described in Section 2.1.2, simulations are also performed to assess whether shear applied during cross-linking induces persistent morphological differences. These simulations utilize the same MPCD parameters and parallel-plate geometry as in the cross-linking runs but with no shear (*i.e.*, stationary plates). Linker reactions are also disabled to prevent further cross-linking. The final configuration from the SCNP-formation period is used as the initial state for these simulations. Each simulation is run for $1.2 \times 10^7$ time steps. The first $2 \times 10^6$ steps are discarded for equilibration/system configurations are saved every $10^6$ steps over the remaining $10^7$ steps for analysis. Across all replicates and parameter combinations, this procedure yields an additional 108 000 SCNP configurations.

## 2.2 Dimensionless quantities

Several dimensionless quantities are defined to characterize systems and help elucidate morphological outcomes. The number of backbone beads $N$ is an important parameter that is naturally dimensionless. The linker fraction describes the density of reactive moieties on the precursor polymer chain and is given by

$$f = \frac{N_\ell}{N}. \tag{4}$$

The distribution or patterning of linkers is described by a normalized blockiness parameter, $\beta$, given by

$$\beta = \frac{b(f) - b_{\min}(f)}{1 - b_{\min}(f)}, \tag{5}$$

with

$$b(f) = \frac{1}{N-1}\sum_{k=0}^{N-2} 1_{k,k+1} \tag{6}$$

and

$$b_{\min}(f) = |2(f - 0.5)|. \tag{7}$$

In eqn (6), the summation is over backbone beads, and $1_{k,k+1}$ is an indicator function equal to unity when the indexed beads are both functionalized (or not) with a linker bead and equal to zero otherwise. The resistance to bending for a polymer is compared to thermal energy through the dimensionless quantity $\frac{\kappa\varepsilon}{k_\mathrm{B}T}$. Fig. S1 provides a relationship between $\frac{\kappa\varepsilon}{k_\mathrm{B}T}$ and the persistence length. The strength of shear flow relative to characteristics of the polymer is captured *via* the Weissenberg number,

$$\mathrm{Wi} = \dot{\gamma}\tau_\mathrm{p}, \tag{8}$$

where $\tau_\mathrm{p}$ describes a polymer relaxation time, which depends on $\kappa$ and $N$.[53] Details of the relaxation time computation are provided in Section 2.3.5. Experimentally, several strategies exist to tune both the degree of polymerization and the stiffness of polymer chains. For instance, incorporation of side-chain groups has been shown to increase backbone rigidity in systems such as polystyrene[54] and polynorbornene,[55] while block copolymerization between side-chain-bearing and unsubstituted polystyrene provides additional control over stiffness.[54] The degree of polymerization can be modulated by adjusting the stoichiometric ratio of monomers to chain transfer agents,[41,43] or by employing physical methods such as sonication.[56] Furthermore, synthetic approaches allow for precise positioning of reactive side groups along the backbone.[25,45,57] Thus, the simulated parameter space corresponds directly to experimentally accessible properties.

**2.2.1 Scope of parameter space.** Precursor polymer chains are constructed according to the specification of the degree of polymerization $N$, linker fraction $f$, normalized blockiness $\beta$, and polymer stiffness $\kappa$. In this work, we investigate precursor chains with $N \in \{50, 100, 150\}, f \in \{0.1, 0.2, 0.4\}$, $\beta \in \{0.2, 0.8\}$, and $\frac{\kappa\varepsilon}{k_\mathrm{B}T} \in \{0, 5, 10\}$. Although we consider precursors with 50–150 beads, each bead may be interpreted as representing multiple monomeric units. This allows our coarse-grained model to capture the behavior of SCNPs at molecular weights more comparable to those reported experimentally.[58–60] The values of $f$ are informed by the work of Patel *et al.*[4] who reported that the effect of $f$ diminished above this range. Moreover, the values of $\beta$ simply span low- and high-blockiness regimes, as its influence was found to be secondary to that given by $f$. The incorporation of $\kappa \neq 0$ is distinct from the work by Patel *et al.*[4] Lastly, shear rates are set as $\dot{\gamma} \in \{0, 0.001, 0.005, 0.01\}$, yielding Weissenberg numbers Wi $\sim 10^0$–$10^2$. The combination ($N, f, \beta, \kappa$, Wi) defines a parameter set that fully specifies the system and conditions for the simulations.

## 2.3 Analysis

**2.3.1 Standard structural characterization.** Simulation trajectories are used to compute the gyration tensor $\mathbf{G} \in \mathbb{R}^{d\times d}$ defined as

$$\mathbf{G} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_\mathrm{cm})(\mathbf{r}_i - \mathbf{r}_\mathrm{cm})^\mathrm{T}, \tag{9}$$

where $\mathbf{r}_i \in \mathbb{R}^3$ is the position vector of bead $i$, and the center of mass is

$$\mathbf{r}_\mathrm{cm} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i. \tag{10}$$

Diagonalization yields $\mathbf{G} = \mathrm{diag}(\lambda_1{}^2, \lambda_2{}^2, \lambda_3{}^2)$ where the diagonal elements $\lambda_1{}^2 \geq \lambda_2{}^2 \geq \lambda_3{}^2$ are the principal moments of the gyration tensor. Subsequently, the radius of gyration is computed as

$$R_g = \sqrt{\lambda_1{}^2 + \lambda_2{}^2 + \lambda_3{}^2}, \tag{11}$$

which generally characterizes the size of a polymer in a given configuration. In addition, the relative shape anisotropy is computed as

$$A_3 = \frac{3}{2}\frac{\lambda_1{}^4 + \lambda_2{}^4 + \lambda_3{}^4}{\left(\lambda_1{}^2 + \lambda_2{}^2 + \lambda_3{}^2\right)^2} - \frac{1}{2}, \tag{12}$$

which provides insight into the distribution of beads rather than size. In particular, $A_3 \in [0, 1]$, with $A_3 = 0$ for an isotropic object and $A_3 = 1$ for a fully extended rod.

The topology of SCNPs is also often considered when characterizing their morphology. Inspired by Moreno *et al.*,[27] we use the distribution of topological domain sizes to assess whether shear applied during cross-linking leads to persistent differences in bonding patterns. Fig. S4 schematically depicts an SCNP with two topological domains. To compute topological domain sizes, we first identify the set of all bonds between linkers, denoted as $\mathscr{B} = \{(a,b)\}$. Each bonded pair $(a,b) \in \mathscr{B}$ is then associated with a contour interval along the polymer backbone $P_{(a,b)}$. Merging overlapping $P_{(a,b)}$ yields $N_{\text{dom}}$ non-overlapping segments; the size, or number of backbone beads, associated with a domain $i$ is denoted $n_i$. This enables calculation of a median domain size $\tilde{n}_i$. For analysis, data are partitioned into three regimes dependent on Wi: small, $0 \le \text{Wi} < 10$; intermediate, $10 \le \text{Wi} < 50$; and large, $\text{Wi} \ge 50$. Within each regime, we further condition on selected combinations of $f$, $\kappa\varepsilon/k_BT$, and $N$ to compare the distributions of $\tilde{n}_i$.

**2.3.2 Unsupervised learning of SCNP morphologies.** Dimensionality reduction by unsupervised learning is used to visualize and characterize SCNP morphologies. For each SCNP configuration, a probability density histogram comprised of all backbone bead pairwise distances is used as a translationally- and rotationally-invariant representation of SCNP structure. This histogram is discretized into 40 evenly-sized bins on the interval $[0.8460\sigma, 140.7\sigma]$ and normalized such that $\sum_{i=1}^{n_{\text{bins}}} (\text{bin values}) \times (\text{bin width}) = 1$, yielding a 40-dimensional vector that describes the SCNP configuration. This 40-dimensional vector is then transformed to a three-dimensional vector using the uniform manifold approximation and projection algorithm (UMAP; umap-learn 0.5.7). The UMAP algorithm[61] first constructs a a weighted nearest-neighbor graph in the initial high-dimensional vector space. Then it optimizes a low-dimensional embedding by minimizing the cross-entropy between the original and embedded edge-weight distributions; this strategy aims to preserve both local and global structure of the neighborhood graph. Compared with *t*-distributed stochastic neighbor embedding (*t*-SNE), UMAP provides comparable performance while being faster.[61] As shown in Fig. S8, both methods yield embeddings with different global shapes but capture consistent morphological trends in the SCNP data.

The UMAP algorithm depends on specification of some hyperparameters. Here, the hyperparameters used are *n*_neighbors = 1000 and min_dist = 1. When *n*_neighbors is large, the optimization emphasizes global structure over local detail. Fig. S7 illustrates the effect of varying these hyperparameters. Although the absolute arrangement of points changes, structural trends remain robust across settings.

The overall procedure defines a mapping $\mathbb{R}^{3N} \rightarrow \mathbb{R}^{40} \rightarrow \mathbb{R}^3$ where the last vector space has axes $\mathbf{Z} = (Z_1, Z_2, Z_3)$. Coordinates in $\mathbf{Z}$ are a learned distillation of the SCNP morphology, which subsequently we aim to interpret by property annotation and other modeling.

**2.3.3 Regression models and feature importance.** Linear regression models are employed to evaluate the utility and relative importance of system parameters in determining SCNP morphology. As a physically meaningful descriptor of morphology, we consider a size-contour ratio for the SCNP, quantified by $R_g/L_c$, where $L_c$ is the polymer contour length. The regression models use the median value of this ratio, $\widetilde{\left(\frac{R_g}{L_c}\right)}$, as the dependent variable. We choose the median as the target for prediction rather than the mean because it is more robust to outliers and is guaranteed to correspond to an observed morphology. Dependent variables include the dimensionless system parameters: $N, f, \beta, \frac{\kappa\varepsilon}{k_BT}$, and Wi. The model representation is constructed using third-order polynomials based on these variables. For comparison, Fig. S9 shows results obtained with a random forest model, which closely align with those of the polynomial regression.

For each model, the median is determined by aggregating results per each unique combination of system parameters. For computing $\frac{R_g}{L_c}$, the contour length is approximated as $L_c = \sum_{i=1}^{N-1} |\mathbf{r}_{i+1} - \mathbf{r}_i|$, where $N$ is the degree of polymerization and $\mathbf{r}_i$ denotes the position of monomer $i$. To assess feature importance, we adopt a permutation-based approach in which each input variable is independently shuffled, the regression model is retrained, and the resulting change in the coefficient of determination $(R^2)$ is recorded. A larger decrease in $R^2$ indicates that the permuted feature plays a more significant role in predicting the median morphology metric $\widetilde{\left(\frac{R_g}{L_c}\right)}$, and thus reflects a stronger influence of the corresponding parameter on SCNP structure. We repeat this analysis using random forest regression and obtain similar results, as shown in Fig. S9.

**2.3.4 Assessment of statistical significance.** To test whether shear applied during cross-linking alters the equilibrium morphology of single-chain nanoparticles, separate hypothesis tests are performed for the number of topological domains, $n_i$, and the distribution of the size-contour ratio $P(R_g/L_c)$. For each descriptor, we fit a simple linear model that relates the observable to the imposed Wi,

$$y_n = c_0 + c_1 \text{ Wi}_n + \nu_n, \tag{13}$$

where $y_n$ is the measured response for sample $n$ and $\mathrm{Wi}_n$ is Wi for the given sample, and $\nu_n \sim \mathcal{N}(0, {\omega_{\mathrm{obs}}}^2)$ is a Gaussian noise term. The intercept $c_0$ and slope $c_1$ are assigned independent uniform priors over physically plausible ranges. Following the recommendations of Gelman,[62] a half-Cauchy prior is placed on the unknown noise scale,

$$p(\omega_{\mathrm{obs}}) = \frac{2}{\pi\omega_0}\frac{1}{1+(\omega_{\mathrm{obs}}/\omega_0)^2}, \tag{14}$$

with $\omega_{\mathrm{obs}} \geq 0$ and the scale parameter $\omega_0$ set to the empirical standard deviation of the observations. The prior is evaluated on a log-spaced grid spanning $0.01\omega_0$ to $10\omega_0$.

To evaluate the posterior, a Cartesian grid is used for $c_0$, $c_1$, and $\omega_{\mathrm{obs}}$. At each grid point $(c_{0,i}, c_{1,j}, \omega_k)$ the Gaussian log-likelihood is

$$\begin{aligned} \log p(\mathbf{y}|c_{0,i}, c_{1,j}, \omega_k) \;=\; & -\frac{N}{2}\log(2\pi{\omega_k}^2) \\ & -\frac{1}{2{\omega_k}^2}\sum_{n=1}^{N}(y_n - c_{0,i} - c_{1,j}\mathrm{Wi}_n)^2. \end{aligned} \tag{15}$$

Adding the log-prior for $\omega_{\mathrm{obs}}$ yields the joint log-posterior. Numerically marginalizing over $\omega_{\mathrm{obs}}$ (summing over $\omega_k$) gives a two-dimensional posterior surface $p(c_0, c_1|\mathbf{y})$, from which samples are drawn of $(c_0, c_1)$. A shear-induced effect is deemed statistically significant at the 5% level if at least 97.5% of the posterior mass for the slope satisfies $c_1 > 0$ or $c_1 < 0$. Equivalently, the 95% Bayesian credible interval for $c_1$ must exclude zero. Rejecting the null hypothesis $c_1 = 0$ indicates that the corresponding morphological property is systematically affected by shear applied during SCNP formation.

**2.3.5 Computation of polymer relaxation time.** To obtain the Weissenberg number, Wi, we must compute a characteristic polymer relaxation time $\tau_{\mathrm{p}}$. For this, we use a representative end-to-end vector relaxation time, $\tau_{\mathbf{R}}(\kappa, N)$, which is a function of $\kappa$ and $N$. For each precursor chain, $i$, we compute the normalized correlation function

$$C_{\mathbf{RR}}^{(i)}(t) = \frac{\langle \mathbf{R}(t'+t)\cdot\mathbf{R}(t')\rangle_i}{\langle \mathbf{R}(t')\cdot\mathbf{R}(t')\rangle_i} \tag{16}$$

where $\mathbf{R}(t) = \mathbf{r}_N(t) - \mathbf{r}_1(t)$ is the end-to-end vector at time $t$, and $\langle\cdot\rangle_i$ indicates an ensemble average for a simulation of the chain $i$, which here is treated equivalently to a time average (*i.e.*, all configurations can be used as a representative at $t'$). The normalized correlation function is then approximately fit as

$$C^{(i)}(t) \approx \exp[-t/\tau_{\mathbf{R}}^{(i)}], \tag{17}$$

to obtain $\tau_{\mathbf{R}}^{(i)}$. Representative autocorrelation decays are shown in Fig. S2. Data is then aggregated across all trajectories containing chains with the same $\kappa$ and $N$ to obtain the median. We employ the relation

$$\tau_{\mathbf{R}}(\kappa, N) = 0.051N\left(\frac{\kappa\varepsilon}{k_{\mathrm{B}}T}N + \frac{\kappa\varepsilon}{k_{\mathrm{B}}T} + N\right), \tag{18}$$

which is obtained using symbolic regression to data on the median end-to-end vector relaxation time observed for precursor chains with a given $\kappa$ and $N$. Eqn (18) is also shown in Fig. S2. This approach produces reasonable scaling behavior to distinguish systems of differing $\kappa$ and $N$.

For this procedure, 540 separate simulations are performed, spanning 54 distinct precursor chain parameter combinations, with 10 unique chains per combination. In these simulations, cross-linking is disabled, and the parameter space is defined as $f \in \{0.1, 0.2, 0.4\}$, $\frac{\kappa\varepsilon}{k_{\mathrm{B}}T} \in \{0, 5, 10\}$, $N \in \{50, 100, 150\}$, and $\beta \in \{0.2, 0.8\}$. Each simulation is run for $5 \times 10^6$ steps, with configurations saved every 100 000 steps.

## 3 Results and discussion

### 3.1 Organization of SCNP morphologies

To broadly characterize the morphologies of the SCNPs, we use unsupervised learning to project each configuration into a three-dimensional embedding space with coordinate axes $\boldsymbol{Z} = (Z_1, Z_2, Z_3)$. Unless otherwise noted, the configurations analyzed here are sampled while the systems remain under the same shear conditions as those applied during cross-linking. In this space, we observe a continuum of morphologies, with similar morphologies positioned in close proximity.

Fig. 2a illustrates that the embedding space is organized primarily by $R_{\mathrm{g}}$. Moving from left (high $Z_1$) to right (low $Z_1$), $R_{\mathrm{g}}$ increases, corresponding to a transition from globular (low $R_{\mathrm{g}}$, blue) to stretched (high $R_{\mathrm{g}}$, yellow) configurations. Additionally, Fig. 2b demonstrates a secondary organization according to the relative shape anisotropy, $A_3$. In particular, we observe a shift from spherical to asymmetrical structures along bands of constant $R_{\mathrm{g}}$ as one moves across each band. The apparent organization by $R_{\mathrm{g}}$ and $A_3$, without explicit supervision on either descriptors, suggests these as major distinctive features of SCNP morphologies.

The low-dimensional embedding arranges SCNP configurations in an overall intuitive manner. Fig. 2c provides specific connection to morphological classes by overlaying representative chain snapshots onto the embedding space. The upper left portion of the plot ($Z_1 \gtrsim 5$) is populated by compact, nearly spherical SCNPs. Traversing the curved manifold rightward toward smaller $Z_1$ yields progressively more elongated structures, illustrating the secondary organization by asphericity. By further decreasing $Z_2$ and $Z_1$, the morphologies evolve into highly anisotropic tadpole and necklace configurations. For example, Fig. 3 displays eight exemplar chains from specific regions of the embedding space. The visual samples are reminiscent of globular (Fig. 3a),[4,27] tadpole (Fig. 3b),[4,63] and necklace-like (Fig. 3c)[4,28] morphologies previously reported in the literature. The figure further shows how locality in the embedding space implies morphological similarity. An increased prevalence of stretched, linear structures likely results from shear applied during simulations.

This data-driven approach may be helpful in revealing hard-to-characterize morphologies or otherwise providing

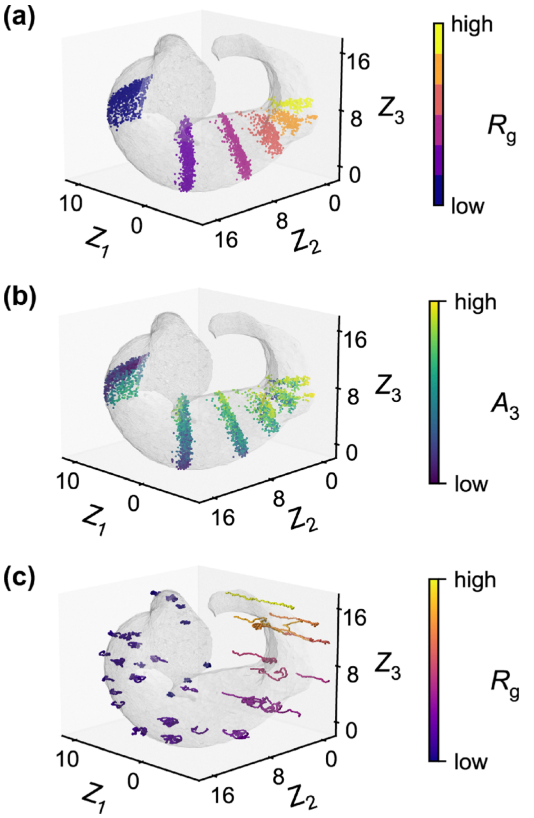


Fig. 2 Organization of morphologies for SCNPs in a three-dimensional embedding space determined by the uniform manifold approximation and projection (UMAP) unsupervised learning algorithm. For clarity, subplots (a) and (b) display slices of the embedding space rather than the entire manifold, with a continuous spectrum of morphologies present between the depicted bands. (a) Six bands of constant $R_g$ across the embedding space. (b) The same bands in (a) but annotated by the relative shape anisotropy. In both (a) and (b), each marker represents an SCNP configuration that is mapped to the coordinates along the embedding axes, $Z_1$, $Z_2$, and $Z_3$. (c) A subset of chain conformations depicted near their coordinates in the embedding space. In each panel, a gray silhouette traces the full extent of the configuration manifold within the embedding space. All quantities are colored by the value associated with single configurations.

quantitative connections between morphologies and their corresponding precursor attributes. For instance, if the goal is to generate tadpole-like structures for which a precise mathematical definition is lacking, one may examine the embedding space (Fig. 3a–c) until a region of desired structures is identified (Fig. 3b). By selecting a small region of the embedding space, observing the predominant morphology, and recording the chain attributes and Weissenberg conditions associated with that region, one can, in principle, design precursor chains that are likely to form the target morphology.

### 3.2 Unsupervised learning reveals property–morphology relationships

To explore how precursor chain variables affect final structures, we map the dominant value of each input feature across the embedding space. Specifically, in Fig. 4, we annotate (by color) each SCNP configuration in the embedding space with system attributes. We also color by $R_g$ to verify the trend between position and $R_g$ noted in Fig. 2a and compare to variation in attributes.

Fig. 4 reveals several broad trends with respect to the studied parameters. The connection between $R_g$ and embedding space location is fairly smooth, shifting from smaller values on one side to larger values on the other (Fig. 4a). Higher $f$ generally produces more compact conformations (Fig. 4b), and $\frac{\kappa\varepsilon}{k_B T}$ tends to correlate with larger $R_g$, although there is considerable overlap between $\frac{\kappa\varepsilon}{k_B T} = 5$ and $\frac{\kappa\varepsilon}{k_B T} = 10$ (Fig. 4c). The effect of $N$ is less pronounced. While the high-$R_g$ region is primarily populated by long chains ($N = 150$), a subset of short chains ($N = 50$) on the right side of the embedding space also exhibits large $R_g$ (Fig. 4d). Cross-referencing between Fig. 4b and c reveals that these high-$R_g$ short chains have low $f$ and high $\frac{\kappa\varepsilon}{k_B T}$, suggesting that linker fraction and bending rigidity can overshadow chain length $N$ in determining morphological outcomes. This is consistent with the wormlike-chain relation proposed by Benoit and Doty,[64,65]

$$R_g^2 = \frac{L_c L_p}{3} - L_p^2 + \frac{2L_p^3}{L_c}\left[1 - \frac{L_p}{L_c}\left(1 - e^{-L_c/L_p}\right)\right], \qquad (19)$$

which shows that for a persistence length $L_p$ and contour length $L_c$, $R_g \propto \sqrt{L_c}$ if $L_c \gg L_p$. Since $L_c \propto N$, a change in

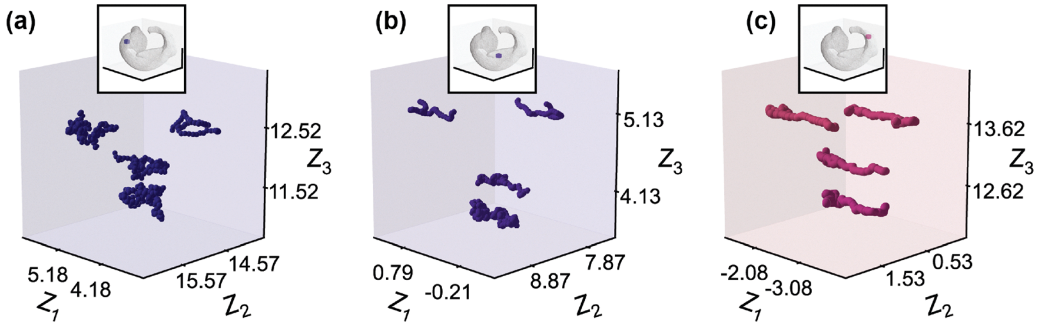


Fig. 3 Representative regions of the embedding space and corresponding SCNP morphologies. (a) Globular chains at low $Z_1$. (b) Semi-elongated structures at intermediate $Z_1$. (c) Elongated rod-like morphologies at high $Z_1$. Insets indicate the location of each zoomed-in region within the embedding space.

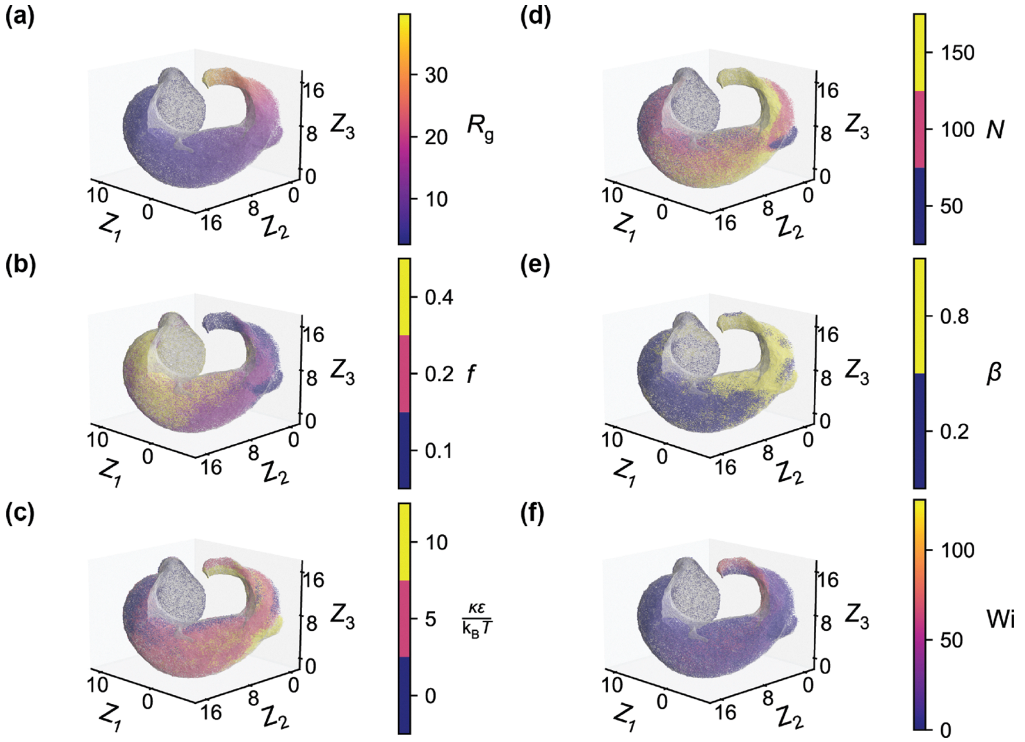


Fig. 4 Correlation between precursor properties (b)–(f) and embedding space regions. Points are colored by the median within a small radius. Panels show the embedding colored by: (a) radius of gyration, $R_g$ (included to illustrate how precursor chain properties map to structural outcomes), (b) linker fraction, $f$, (c) reduced stiffness, $\frac{\kappa\varepsilon}{k_BT}$, (d) degree of polymerization, $N$, (e) relative linker blockiness, $\beta$, and (f) Weissenberg number, Wi.

$N$ is expected to yield a comparably small change in $R_g$ as $N$ becomes large.

Beyond these broad trends, the arrangement of linker beads along the backbone plays a smaller role. Examining the effect of $\beta$ (Fig. 4e) reveals that regions with higher $R_g$ often display blockier arrangements, but numerous exceptions imply that $\beta$ has less influence than other features. Finally, when SCNPs are under shear, the Weissenberg number (Fig. 4f) shows a weak positive relationship with $R_g$, but the visual correlation is less pronounced compared to other inputs, particularly $f$ or $\frac{\kappa\varepsilon}{k_BT}$. However, because shear is an external condition rather than a parameter determined by precursor chemistry, it merits consideration as a means to bias morphological outcomes[31,66,67] when precursor chemistry is constrained.

### 3.3 Analysis of individual parameter effects on morphology

From Fig. 4a–c, $f$ and $\frac{\kappa\varepsilon}{k_BT}$ appear strongly correlated with $R_g$. Motivated by this qualitative observation, we aim to more clearly assess influence by examining the marginal distribution of the size-contour ratio, $\frac{R_g}{L_c}$, with respect to $f$ and $\frac{\kappa\varepsilon}{k_BT}$. The marginal distribution of a variable describes its overall variability and central tendency by averaging over other variables, thereby conveying the isolated behavior of that variable irrespective of any dependencies.

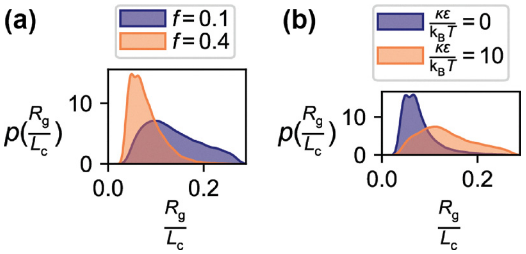


Fig. 5 Marginal distributions of the dimensionless radius of gyration, $\frac{R_g}{L_c}$, for SCNPs with (a) linker fraction $f = \{0.1, 0.4\}$ and (b) backbone stiffness $\frac{\kappa\varepsilon}{k_BT} = \{0, 10\}$.

Fig. 5 shows that as the linker fraction increases, the distribution shifts toward smaller $\frac{R_g}{L_c}$ values and narrows, reflecting more compact configurations (Fig. 5a); this behavior aligns with previous findings by Patel *et al.*[4] In contrast, increasing chain stiffness $\left(\frac{\kappa\varepsilon}{k_BT}\right)$ pushes the distribution toward larger values and broadens it, indicating that stiffer chains tend to remain more extended (Fig. 5b). This result contrasts with the findings of Moreno *et al.*, who observed that stiffer chains formed more compact structures;[27] however, Moreno *et al.* based cross-link formation solely on a distance criterion between linker beads, whereas we adopt both distance and angle-based criteria.[4]

We speculate that this discrepancy arises from our inclusion of an angle criterion during bond formation. In the model of Moreno *et al.*, cross-links form solely based on the spatial proximity of linker beads, with the additional restriction that linkers bound to directly adjacent backbone beads cannot react with one another. However, this still permits reactions between linkers attached to backbone beads separated by only one monomer (*i.e.*, neighbor + 1 linkers). Among flexible chains, such short-range reactions rapidly deplete available linkers, thereby hindering global chain collapse. Steric constraints in stiff chains inhibit these local reactions, permitting the formation of longer-range cross-links. By contrast, our model imposes both distance- and angle-based criteria for bond formation. This combined requirement suppresses near-neighbor reactions irrespective of backbone stiffness, thereby enabling long-range cross-linking in flexible chains. Fig. S10a shows that disabling the angle criterion, as in the approach of Moreno *et al.*, yields more extended morphologies compared to those we obtained.

Next, we quantify the relative importance of each input feature in determining the fully cross-linked morphology. To do so, we fit a third-order polynomial, using system parameters as dependent variables, to predict the median $\frac{R_g}{L_c}$, $\left(\widetilde{\frac{R_g}{L_c}}\right)$. We then assess feature importance by randomly shuffling each variable one at a time, refitting the model, and recording the change in the coefficient of determination $R^2$ (feature permutation). This process is repeated 1000 times per feature to ensure robustness; variables that cause a larger change in $R^2$ are considered more influential.

Fig. 6a shows that this model achieves strong predictive performance. This suggests that the dimensionless quantities are meaningfully connected to the size-contour ratio, and that evaluating the importance of these features for model predictions may be more likely connected to the underlying physical trends as well. Following this, Fig. 6b shows that $f$ has the greatest impact, while $\beta$ has the least influence under the simulated conditions. These results are consistent with findings from Patel *et al.*[4] who reported that linker fraction has a stronger influence on morphological outcomes than blockiness. Notably, the Weissenberg number, which depends on relaxation time (Fig. S2) and shear rate, plays a greater role in determining morphology than blockiness. This is despite the latter being an intrinsic property of the chain and shear rate being an externally imposed condition.

To visualize the combined effect of the three most influential variables on $\left(\widetilde{\frac{R_g}{L_c}}\right)$, we consider the design space spanned by those features in Fig. 6c, which conveys a clear trend. Namely, $\left(\widetilde{\frac{R_g}{L_c}}\right)$ decreases (color change from yellow/orange towards red/purple) when transitioning from short, stiff chains with low linker fraction (bottom left) to long, flexible chains with higher linker fraction (upper right). This confirms these parameters as dominant determinants of morphology. In addition, decreasing $f$ or increasing $\kappa$ tends to increase variability (increasing size of spheres). The implication is that flexible precursor chains with a high fraction of crosslinking moieties consistently produce compact morphologies, whereas chains with either high stiffness or a low linker fraction yield SCNPs with more variable structures. These trends are similar to those found in morphology dispersity measures proposed by Patel *et al.*[4] but now observed for SCNPs formed under shear flow.

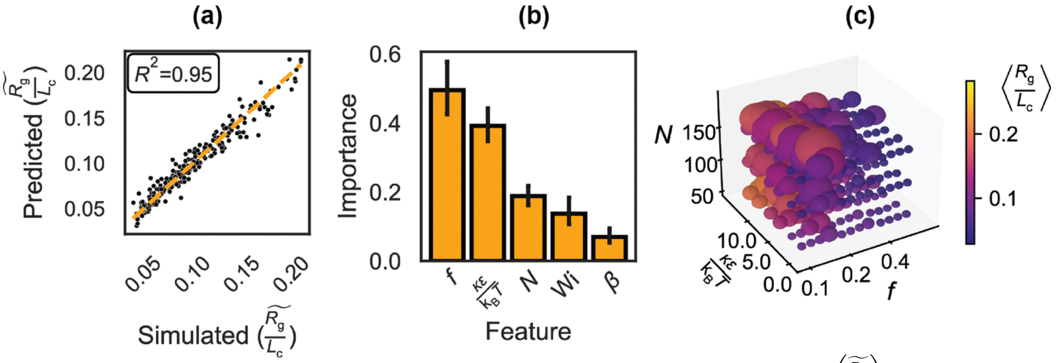


**Fig. 6** Analysis of importance of input parameters based on polynomial regression. (a) Predicted *versus* actual $\left(\widetilde{\frac{R_g}{L_c}}\right)$. Each data point corresponds to the set of chains that share the same standard parameter space values. (b) Parameter importance by feature permutation. The data provides the median $R^2$ change from 1000 random shuffles per variable, with error bars indicating the interquartile range. A larger value indicates greater importance. (c) Three-dimensional scatter of the top three predictors $f$, $\frac{\kappa\varepsilon}{k_BT}$, and $N$, in which the color of each sphere denotes $\left(\widetilde{\frac{R_g}{L_c}}\right)$ and its radius corresponds to the interquartile range. $\left(\widetilde{\frac{R_g}{L_c}}\right)$ is computed for each point in the standard parameter space, and spheres are spaced slightly to separate each ($\beta$, Wi) combination associated with the same $\left(f, \frac{\kappa\varepsilon}{k_BT}, N\right)$.

### 3.4 Influence of chain characteristics on shear-induced morphological changes

We next demonstrate how the effect of shear on morphology depends on precursor chain characteristics by specifically contrasting systems formed with Wi = 0 *versus* Wi > 0. Fig. 7a and b illustrates cases where shear increases or decreases $\left(\widetilde{\frac{R_g}{L_c}}\right)$, respectively. In the first case, which relates to flexible precursor chains with a low linker fraction, shear leads to more extended conformations (Fig. 7c and d). Under quiescent conditions (Wi = 0), linker beads can more easily react due to the flexibility of the chain. However, under flow, shear stress exerts a force on the chain, stretching it into an elongated conformation rarely observed under quiescent conditions. This alignment reduces the likelihood of cross-linking by increasing the total distance between reactive sites while shear is maintained.

In the second case, which relates to precursor chains with high stiffness and a high linker fraction, shear promotes more globular structures (Fig. 7e and f). We speculate that this effect results from the competition between shear stress and backbone rigidity. Without shear, stiffness inhibits reactive moieties from coming into proximity, even at high linker fractions. However, under shear, the velocity gradient induces tumbling and transient bending as the chain aligns with the flow direction. This shear-induced flexing occurs because, at high Wi, the applied stress overcomes the bending potential of the chain, drawing reactive sites closer together and increasing the number of cross-links. Consequently, we observe more compact structures relative to similar chains under quiescent

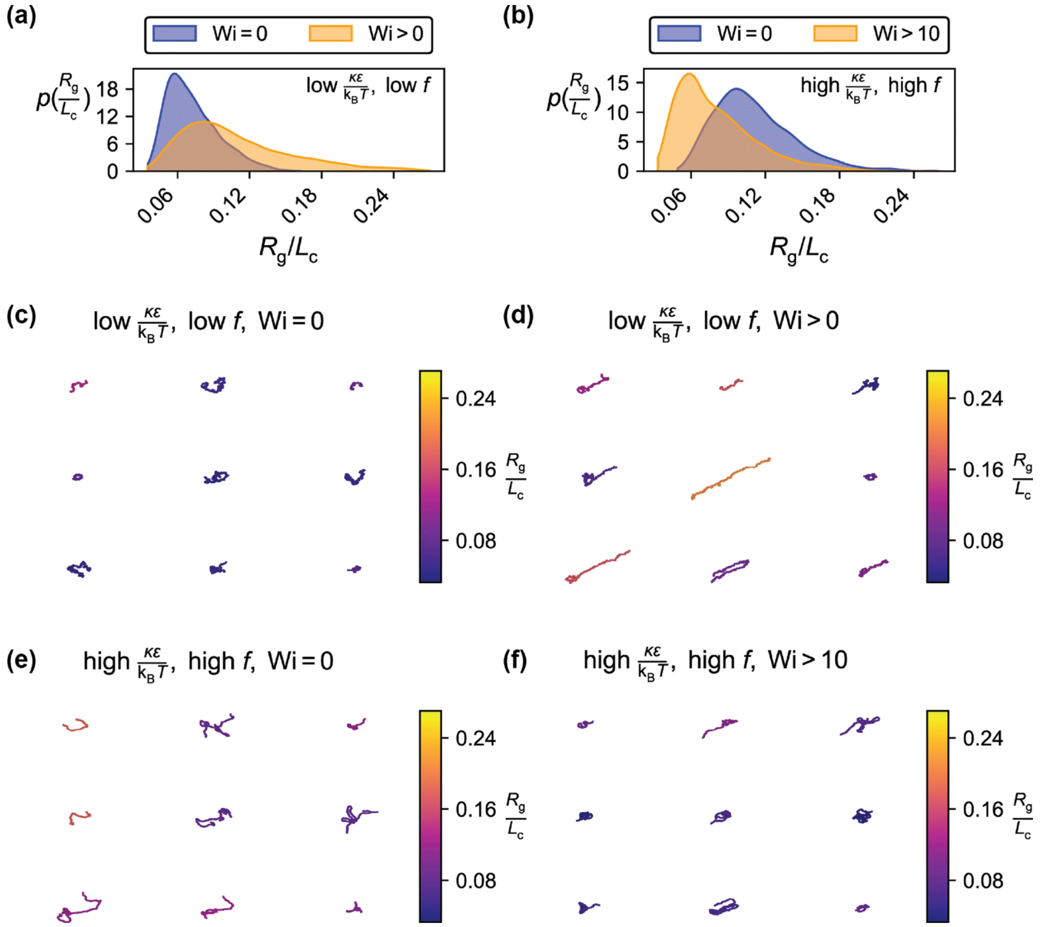


Fig. 7 Analysis of the effect of shear on SCNPs. Observed distributions of $\frac{R_g}{L_c}$ for SCNPs formed under quiescent conditions (Wi = 0) and in shear flow (Wi > 0) for precursor chains with (a) flexible chains with few reactive linkers $\left(\frac{\kappa\varepsilon}{k_BT} = 0, f = 0.1\right)$ and (b) stiff chains with a high density of reactive linkers $\left(\frac{\kappa\varepsilon}{k_BT} = 10, f = 0.4\right)$. In (a), nonzero Wi raises the median relative to quiescent conditions; in (b), nonzero Wi lowers the median. Representative configurations of SCNPs with low $\left(\frac{\kappa\varepsilon}{k_BT}, f\right)$ (*i.e.*, the same attributes as in panel a) under (c) no shear and (d) shear conditions. These SCNPs tend to be globular under quiescent conditions but stretch under shear. Representative configurations of SCNPs with high $\left(\frac{\kappa\varepsilon}{k_BT}, f\right)$ (*i.e.*, the same attributes as in panel b) under (e) no shear and (f) shear conditions. These SCNPs tend to exhibit tadpole and necklace-like shapes at Wi = 0 but become more compact at elevated Wi.

conditions. This counterintuitive trend is consistent with the mean-field model of Winkler for semiflexible polymers in shear, which predicts that increasing shear rate causes semiflexible chains to behave more like flexible chains.[68] Experimental studies on actin filaments in shear further support this interpretation, as they directly reveal buckling, high-curvature U-turns, and snaking motions that transiently bring backbone segments separated by large contour distances into contact.[35,69] Thus, even though shear is not the dominant factor controlling morphology, it can nonetheless bias outcomes across precursor chains with shared attributes. While synthesis of SCNPs under shear flow has not yet been systematically explored, such conditions are naturally present in many polymer processing environments. Our simulations therefore provide a conceptual framework for how shear could influence SCNP formation and open avenues for future experimental investigation.

### 3.5 Persistent effects of non-equilibrium cross-linking

To determine whether shear applied during cross-linking has a persistent effect on SCNP topology, we examine how $f$, $\frac{\kappa\varepsilon}{k_BT}$, and Wi affect the distribution of topological domain sizes within the SCNP. A topological domain in an SCNP comprises all monomeric units between the two most distant reacted linkers. Consequently, SCNPs with only a few topological domains (and large domain sizes) usually possess highly crosslinked, dense network structures, whereas those with many domains (and lesser domain sizes) exhibit more fragmented and loosely connected architectures.

Fig. 8a shows that for all $\frac{\kappa\varepsilon}{k_BT} \geq 5$ for $N = 150$, increasing Wi shifts the distribution toward larger domain sizes; equivalent analyses for $\frac{\kappa\varepsilon}{k_BT} = 10$ are provided in the SI, Fig. S5. We speculate that general increase in topological domain size at elevated Wi occurs because shear facilitates contacts between contour-distant moieties. Accordingly, applying shear during cross-linking may promote larger loop formation in systems with high backbone rigidity. Moreover, we find that under both low and high Wi, increasing $f$ shifts the distribution toward larger domain sizes. We also find that for $f = 0.4$, increasing $\frac{\kappa\varepsilon}{k_BT}$ increases the median topological domain size, consistent with Moreno *et al.*[27] These results collectively highlight shear flow as a valuable tool for systematically tuning topological connectivity for precursor chains with high stiffness or high linker fraction.

We next evaluate whether shear applied during cross-linking has a persistent effect on SCNP morphology. To do so, we first extract the SCNP configurations obtained under shear, then continue the simulations under quiescent conditions until the structures equilibrate. The relaxed morphologies are subsequently projected onto the previously constructed embedding space, which includes configurations from both Wi = 0 and Wi > 0. Then, we track the difference in centroid coordinates for SCNPs cross-linked under zero-shear *versus* high-shear conditions.

Fig. 8b shows that, for chains with identical $\left(f, \frac{\kappa\varepsilon}{k_BT}, N, \beta\right)$, cross-linking under shear and then relaxing under quiescent conditions yields average morphologies that are distinct from

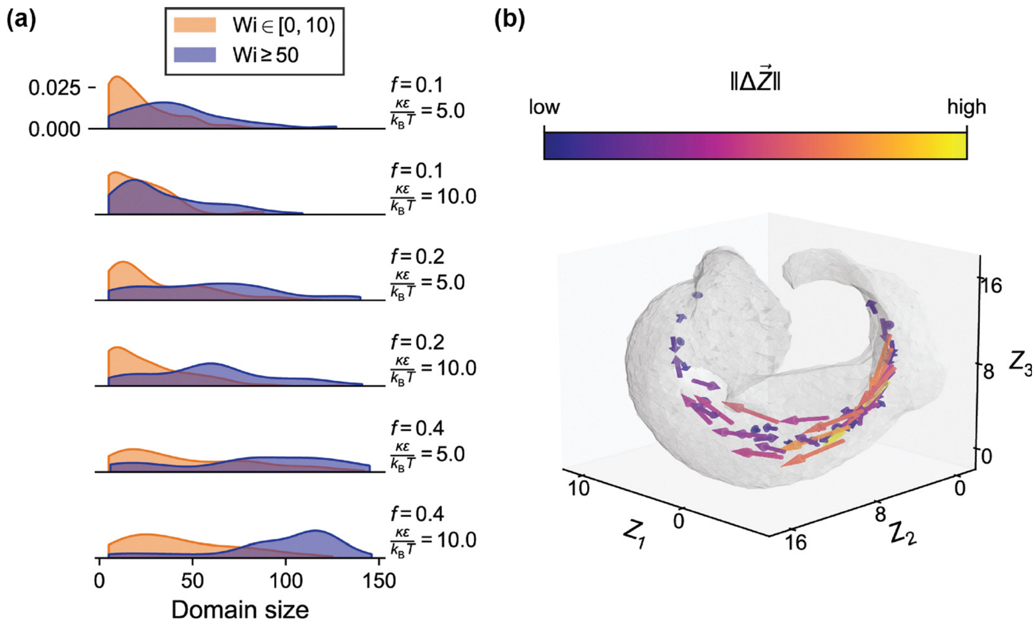


Fig. 8 Examination of shear-induced changes to morphology at quiescent conditions. (a) Observed distributions of SCNP topological domain sizes for selected linker fraction $f$ and stiffness parameters $\frac{\kappa\varepsilon}{k_BT}$; all chains have $N = 150$ beads. Low Weissenberg number (Wi < 10) is shown in orange; high (Wi ≥ 50) in blue. *Y*-Axis bounds are consistent across all plots. (b) Representative sampling of displacement of embedding-space centroids between SCNPs cross-linked at Wi = 0 and at the maximum simulated Wi for each parameter set. Arrows originate at the Wi = 0 centroids and point to the high-Wi centroids. Color encodes the displacement magnitude as another guide to the eye.

those obtained by cross-linking entirely under quiescent conditions. These effects are confirmed to be statistically significant over particular regimes of precursor-chain parameters in the SI, Fig. S6 and accompanying analysis. Overall, SCNPs cross-linked under high-shear conditions compared to chains with the same $\left(f, \frac{\kappa\varepsilon}{k_{\mathrm{B}}T}, N, \beta\right)$ attributes cross-linked at Wi = 0 display a general shift toward higher $Z_1$. Cross-referencing with Fig. 2a indicates that this shift corresponds to an increase in globular character, suggesting that chains cross-linked under shear are more likely to adopt compact morphologies that persist after shear is removed. However, there are also other secondary effects, such as decreases in asphericity, $A_3$ (Fig. 2b). Together, these results show that shear imposed during cross-linking leaves a lasting morphological "memory," biasing the equilibrium ensemble toward more globular, less aspherical SCNPs even after the flow is removed. We attribute this effect to the irreversibility of cross-linking events in our model, which lock in local configurations when two linkers come into proximity. Under shear, the probability of encounters between linkers at large contour distances is increased, thereby biasing the resulting morphology toward collapsed structures.

## 4 Conclusions

We used molecular dynamics simulations and unsupervised machine learning to examine how dimensionless precursor-chain parameters and shear conditions govern SCNP formation. Specifically, we simulated 10 800 unique SCNPs spanning broad parameter spaces defined by five dimensionless variables. Using unsupervised learning, we organized the observed SCNP morphologies into three-dimensional embedding space. This embedding space was primarily distinguished by $R_g$, with secondary organization given by relative shape anisotropy. By inspection, this learned embedding reasonably organized structures into regions corresponding to morphological archetypes—including globular, tadpole-like, necklace-like, and linear conformations. This illustrated general utility of unsupervised learning to capture morphological subtleties without prior specification or expectations on key parameters.

Analysis of embedding space positions identified links between morphological outcomes and precursor chain attributes as well as shear flow. Many observations resonated with prior literature. For example, higher linker fractions consistently produced more compact morphologies, while increased backbone stiffness generally resulted in more extended structures. Interestingly, although the degree of polymerization influenced SCNP size, we observed that stiffness and linker fraction were frequently more dominant than length effects. The influence of linker blockiness was comparatively small. Polynomial regression with permutation-based feature importance analysis quantitatively confirmed these trends. As a more distinctive conceptual contribution, our analysis illustrated that shear, expressed through the Weissenberg number, also had moderate influence. This is notable given that shear can externally imposed and does not depend on precursor chemistry or synthesis, which may be difficult to control.

Additionally, Weissenberg conditions exhibited context-dependent effects. The presence of shear flow biased flexible chains with low linker fractions from compact equilibrium structures to elongated conformations. Conversely, stiff chains with higher linker fractions displayed shear-induced compaction, as flow facilitated bending, thereby increasing reactive bead encounters. These observations underscore that shear not only influences SCNP structures but does so differently depending on the precursor chain attributes. In comparison to SCNPs formed under quiescent conditions, we further found that that applying shear during cross-linking generally biases SCNP formation toward more compact morphologies that persist after shear is removed.

Collectively, our results establish how key dimensionless variables related to precursor-chain attributes and shear conditions jointly influence SCNP morphology. This provides a rational framework for SCNP design. To experimentally validate the trends observed in our simulations, future work could employ small-angle scattering techniques on SCNPs synthesized under comparable Weissenberg conditions. Previous studies have demonstrated the use of small-angle neutron scattering (SANS)[16,59,70,71] and small-angle X-ray scattering (SAXS)[13,72] to characterize SCNP morphology and compare experimental structures with simulation. These methods could therefore be applied to covalently cross-linked SCNPs formed under flow to assess our predictions. Agreement between experiment and simulation would lend support to the proposed mechanisms (*e.g.*, shear-induced compaction of stiff SCNPs with high linker fractions or elongation of flexible SCNPs with low linker fractions) and motivate further study of SCNP formation under shear.

In the future, it may be interesting to assess how these findings generalize (or not) to self-associating filaments, such as single-stranded DNA, or how they can translate to chemically specific systems. Moreover, because our study is confined to covalent cross-linking, exploring how SCNPs assembled *via* reversible bonding respond to shear warrants further investigation. Addressing the first question experimentally and the second *via* coarse-grained MD simulations will deepen our understanding of SCNP formation and inform polymeric material design. Finally, based on some of our results, we hypothesize that morphological outcomes may be sensitive to details of the coarse-grained model. Benchmarking against atomistically detailed systems will help clarify these effects and guide practical modeling approaches for meaningful comparison with experiment.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information:

relation of polymer stiffness parameter $\kappa$ to persistence length; computation of relaxation time; validation of reaction protocol; topological domain sizes; and relaxation under quiescent conditions. See DOI: **https://doi.org/10.1039/d5sm00729a**.

The data associated with this study is publicly accessible as a dataset, SCNP-Shearflow-10k, deposited in the Zenodo data based under accession code **https://doi.org/10.5281/zenodo.17203738**. Relevant code and simulation files associated with this study are available at **https://github.com/webbtheosim/md-simulation-files/tree/main/2025-scnp-shear**.

## Acknowledgements

The authors acknowledge support from the National Science Foundation. The investigation of controlled structure formation in polymers is supported by the National Science Foundation under Grant No. 2237470 (M. A. W.). H. A. S. is partially supported by the National Science Foundation through the Princeton University (PCCM) Materials Research Science and Engineering Center (DMR-2011750) and Grant No. CBET-2246791. 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. These resources include a GPU-based computing cluster purchased with support from the National Science Foundation (Grant No. NSF-MRI: OAC-2320649) (M. A. W.). M. D. C. acknowledges Roshan A. Patel (Princeton University) for helpful discussions.

## Notes and references

1 T. Jin, C. W. Coley and A. Alexander-Katz, *Nat. Chem.*, 2025, **17**, 997–1004.
2 M. Formanek and A. J. Moreno, *Soft Matter*, 2021, **17**, 2223–2233.
3 A. J. Moreno, F. L. Verso, A. Sanchez-Sanchez, A. Arbe, J. Colmenero and J. A. Pomposo, *Macromolecules*, 2013, **46**, 9748–9759.
4 R. A. Patel, S. Colmenares and M. A. Webb, *ACS Polym. Au*, 2023, **3**, 284–294.
5 S. Mavila, O. Eivgi, I. Berkovich and N. G. Lemcoff, *Chem. Rev.*, 2015, **116**, 878–961.
6 C. K. Lyon, A. Prasher, A. M. Hanlon, B. T. Tuten, C. A. Tooley, P. G. Frank and E. B. Berda, *Polym. Chem.*, 2015, **6**, 181–197.
7 T. Terashima, T. Mes, T. F. A. De Greef, M. A. J. Gillissen, P. Besenius, A. R. A. Palmans and E. W. Meijer, *J. Am. Chem. Soc.*, 2011, **133**, 4742–4745.
8 G. Wulff, B. Chong and U. Kolb, *Angew. Chem., Int. Ed.*, 2006, **45**, 2955–2958.
9 H. Rothfuss, N. D. Knöfel, P. W. Roesky and C. Barner-Kowollik, *J. Am. Chem. Soc.*, 2018, **140**, 5875–5881.
10 A. Sanchez-Sanchez, A. Arbe, J. Colmenero and J. A. Pomposo, *ACS Macro Lett.*, 2014, **3**, 439–443.
11 S. Thanneeru, J. K. Nganga, A. S. Amin, B. Liu, L. Jin, A. M. Angeles-Boza and J. He, *ChemCatChem*, 2017, **9**, 1157–1162.
12 J.-H. Ryu, R. T. Chacko, S. Jiwpanich, S. Bickerton, R. P. Babu and S. Thayumanavan, *J. Am. Chem. Soc.*, 2010, **132**, 17227–17235.
13 A. Sanchez-Sanchez, S. Akbari, A. J. Moreno, F. L. Verso, A. Arbe, J. Colmenero and J. A. Pomposo, *Macromol. Rapid Commun.*, 2013, **34**, 1681–1686.
14 A. P. P. Kröger, N. M. Hamelmann, A. Juan, S. Lindhoud and J. M. J. Paulusse, *ACS Appl. Mater. Interfaces*, 2018, **10**, 30946–30951.
15 A. B. Benito, M. K. Aiertza, M. Marradi, L. Gil-Iceta, T. Shekhter Zahavi, B. Szczupak, M. Jiménez-González, T. Reese, E. Scanziani, L. Passoni, M. Matteoli, M. De Maglie, A. Orenstein, M. Oron-Herman, G. Kostenich, L. Buzhansky, E. Gazit, H.-J. Grande, V. Gómez-Vallejo, J. Llop and I. Loinaz, *Biomacromolecules*, 2016, **17**, 3213–3221.
16 I. Perez-Baena, F. Barroso-Bujans, U. Gasser, A. Arbe, A. J. Moreno, J. Colmenero and J. A. Pomposo, *ACS Macro Lett.*, 2013, **2**, 775–779.
17 Y. Liu, S. Pujals, P. J. M. Stals, T. Paulöhrl, S. I. Presolski, E. W. Meijer, L. Albertazzi and A. R. A. Palmans, *J. Am. Chem. Soc.*, 2018, **140**, 3423–3433.
18 J. Rubio-Cervilla, E. González and J. Pomposo, *Nanomaterials*, 2017, **7**, 341.
19 R. Chen and E. B. Berda, *ACS Macro Lett.*, 2020, **9**, 1836–1843.
20 A. M. Hanlon, C. K. Lyon and E. B. Berda, *Macromolecules*, 2015, **49**, 2–14.
21 M. Gonzalez-Burgos, A. Latorre-Sanchez and J. A. Pomposo, *Chem. Soc. Rev.*, 2015, **44**, 6122–6142.
22 J. A. Pomposo, I. Perez-Baena, F. L. Verso, A. J. Moreno, A. Arbe and J. Colmenero, *ACS Macro Lett.*, 2014, **3**, 767–772.
23 Y. Vo, R. Raveendran, C. Cao, L. Tian, R. Y. Lai and M. H. Stenzel, *J. Mater. Chem. B*, 2024, **12**, 12627–12640.
24 S. Liao, L. Wei, L. A. Abriata and F. Stellacci, *Macromolecules*, 2021, **54**, 11459–11467.
25 K. Kyoda, T. Yamamoto and Y. Tezuka, *J. Am. Chem. Soc.*, 2019, **141**, 7526–7536.
26 M. Formanek and A. J. Moreno, *Soft Matter*, 2017, **13**, 6430–6438.
27 A. J. Moreno, P. Bacova, F. L. Verso, A. Arbe, J. Colmenero and J. A. Pomposo, *J. Phys.: Condens. Matter*, 2017, **30**, 034001.
28 F. Lo Verso, J. A. Pomposo, J. Colmenero and A. J. Moreno, *Soft Matter*, 2014, **10**, 4813–4821.
29 F. Lo Verso, J. A. Pomposo, J. Colmenero and A. J. Moreno, *Soft Matter*, 2015, **11**, 1369–1375.
30 J. W. Liu, M. E. Mackay and P. M. Duxbury, *EPL*, 2008, **84**, 46001.
31 M. Formanek and A. J. Moreno, *Macromolecules*, 2019, **52**, 1821–1831.
32 P. Englebienne, P. A. J. Hilbers, E. W. Meijer, T. F. A. De Greef and A. J. Markvoort, *Soft Matter*, 2012, **8**, 7610.
33 E. Mitsoulis and S. G. Hatzikiriakos, *Polymers*, 2021, **13**, 489.

34 S. W. Schneider, S. Nuschele, A. Wixforth, C. Gorzelanny, A. Alexander-Katz, R. R. Netz and M. F. Schneider, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7899–7903.

35 Y. Liu, B. Chakrabarti, D. Saintillan, A. Lindner and O. du Roure, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 9438–9443.

36 T. A. Meyer, C. Ramirez, M. J. Tamasi and A. J. Gormley, *ACS Polym. Au*, 2022, **3**, 141–157.

37 M. A. Webb, N. E. Jackson, P. S. Gil and J. J. de Pablo, *Sci. Adv.*, 2020, **6**, eabc6216.

38 D. Bhattacharya, D. C. Kleeblatt, A. Statt and W. F. Reinhart, *Soft Matter*, 2022, **18**, 5037–5051.

39 A. Statt, D. C. Kleeblatt and W. F. Reinhart, *Soft Matter*, 2021, **17**, 7697–7707.

40 A. Gardin, C. Perego, G. Doni and G. M. Pavan, *Commun. Chem.*, 2022, **5**, 82.

41 G. Gody, T. Maschmeyer, P. B. Zetterlund and S. Perrier, *Nat. Commun.*, 2013, **4**, 2505.

42 J.-F. Lutz, M. Ouchi, D. R. Liu and M. Sawamoto, *Science*, 2013, **341**, 1238149.

43 A. Anastasaki, V. Nikolaou, G. S. Pappas, Q. Zhang, C. Wan, P. Wilson, T. P. Davis, M. R. Whittaker and D. M. Haddleton, *Chem. Sci.*, 2014, **5**, 3536–3542.

44 B. V. K. J. Schmidt, N. Fechler, J. Falkenhagen and J.-F. Lutz, *Nat. Chem.*, 2011, **3**, 234–238.

45 R. K. Roy and J.-F. Lutz, *J. Am. Chem. Soc.*, 2014, **136**, 12888–12891.

46 K. Kremer and G. S. Grest, *J. Chem. Phys.*, 1990, **92**, 5057–5086.

47 A. Malevanets and R. Kapral, *J. Chem. Phys.*, 1999, **110**, 8605–8613.

48 M. P. Howard, A. Z. Panagiotopoulos and A. Nikoubashman, *Comput. Phys. Commun.*, 2018, **230**, 10–20.

49 I. O. Götze, H. Noguchi and G. Gompper, *Phys. Rev. E:Stat., Nonlinear, Soft Matter Phys.*, 2007, **76**, 046705.

50 J. A. Anderson, J. Glaser and S. C. Glotzer, *Comput. Mater. Sci.*, 2020, **173**, 109363.

51 A. Statt, M. P. Howard and A. Z. Panagiotopoulos, *Phys. Rev. Fluids*, 2019, **4**, 043905.

52 Y. M. Wani, P. G. Kovakas, A. Nikoubashman and M. P. Howard, *J. Chem. Phys.*, 2022, **156**, 024901.

53 M. Rubinstein and R. Colby, *Polymer Physics*, Oxford University Press, Oxford New York, 2003.

54 K. Bornani, X. Wang, J. L. Davis, X. Wang, W. Wang, J. P. Hinestrosa, J. W. Mays and S. M. Kilbey II, *Soft Matter*, 2015, **11**, 6509–6519.

55 D. F. Sunday, A. Chremos, T. B. Martin, A. B. Chang, A. B. Burns and R. H. Grubbs, *Macromolecules*, 2020, **53**, 7132–7140.

56 S. Kim, J. Oh, J. Park, B. Lee, T. L. H. Mai, Z. Sun, S. Jeong, Y. Cho, W. Kim and C. Yang, *Angew. Chem., Int. Ed.*, 2024, **63**, e202401097.

57 S. Pfeifer and J.-F. Lutz, *J. Am. Chem. Soc.*, 2007, **129**, 9542–9543.

58 R. Upadhya, N. S. Murthy, C. L. Hoop, S. Kosuri, V. Nanda, J. Kohn, J. Baum and A. J. Gormley, *Macromolecules*, 2019, **52**, 8295–8304.

59 J. Engelke, B. T. Tuten, R. Schweins, H. Komber, L. Barner, L. Plüschke, C. Barner-Kowollik and A. Lederer, *Polym. Chem.*, 2020, **11**, 6559–6578.

60 J. Engelke, S. Boye, B. T. Tuten, L. Barner, C. Barner-Kowollik and A. Lederer, *ACS Macro Lett.*, 2020, **9**, 1569–1575.

61 L. McInnes, J. Healy, N. Saul and L. Großberger, *J. Open Source Software*, 2018, **3**, 861.

62 A. Gelman, *Bayesian Anal.*, 2006, **1**, 515–534.

63 W. Wen, T. Huang, S. Guan, Y. Zhao and A. Chen, *Macromolecules*, 2019, **52**, 2956–2964.

64 H. Benoit and P. Doty, *J. Phys. Chem.*, 1953, **57**, 958–963.

65 A. M. Rosales, H. K. Murnen, S. R. Kline, R. N. Zuckermann and R. A. Segalman, *Soft Matter*, 2012, **8**, 3673.

66 D. Panja, G. T. Barkema and J. M. J. van Leeuwen, *Phys. Rev. E*, 2016, **93**, 042501.

67 P. S. Doyle, B. Ladoux and J.-L. Viovy, *Phys. Rev. Lett.*, 2000, **84**, 4769–4772.

68 R. G. Winkler, *J. Chem. Phys.*, 2010, **133**, 164905.

69 M. Harasim, B. Wunderlich, O. Peleg, M. Kröger and A. R. Bausch, *Phys. Rev. Lett.*, 2013, **110**, 108302.

70 M. González-Burgos, A. Arbe, A. J. Moreno, J. A. Pomposo, A. Radulescu and J. Colmenero, *Macromolecules*, 2018, **51**, 1573–1585.

71 A. Sanchez-Sanchez, S. Akbari, A. Etxeberria, A. Arbe, U. Gasser, A. J. Moreno, J. Colmenero and J. A. Pomposo, *ACS Macro Lett.*, 2013, **2**, 491–495.

72 S. Wijker, D. Dellemme, L. Deng, B. Fehér, I. K. Voets, M. Surin and A. R. A. Palmans, *ACS Macro Lett.*, 2025, **14**, 428–433.