

MARTINI-Compatible Coarse-Grained Model for the Mesoscale Simulation of Peptoids

Mingfei Zhao, Janani Sampath, Sarah Alamdari, Gillian Shen, Chun-Long Chen, Christopher J. Mundy, Jim Pfaendtner, and Andrew L. Ferguson*

Cite This: *J. Phys. Chem. B* 2020, 124, 7745–7764

Read Online

ACCESS |



Metrics & More

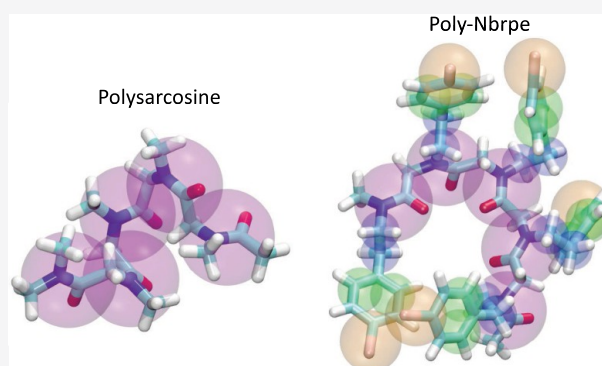


Article Recommendations



Supporting Information

ABSTRACT: Peptoids (poly-N-substituted glycines) are a class of synthetic polymers that are regioisomers of peptides (poly-C-substituted glycines), in which the point of side-chain connectivity is shifted from the backbone C to the N atom. Peptoids have found diverse applications as peptidomimetic drugs, protein mimetic polymers, surfactants, and catalysts. Computational modeling is valuable in the understanding and design of peptoid-based nanomaterials. In this work, we report the bottom-up parameterization of coarse-grained peptoid force fields based on the MARTINI peptide force field against all-atom peptoid simulation data. Our parameterization pipeline iteratively refits coarse-grained bonded interactions using iterative Boltzmann inversion and nonbonded interactions by matching the potential of mean force for chain extension. We assure good sampling of the amide bond cis/trans isomerizations in the all-atom simulation data using parallel bias metadynamics. We develop coarse-grained models for two representative peptoids—polysarcosine (poly(*N*-methyl glycine)) and poly(*N*-((4-bromophenyl)ethyl)glycine)—and show their structural and thermodynamic properties to be in excellent accord with all-atom calculations but up to 25-fold more efficient and compatible with MARTINI force fields. This work establishes a new rigorously parameterized coarse-grained peptoid force field for the understanding and design of peptoid nanomaterials at length and time scales inaccessible to all-atom calculations.



1. INTRODUCTION

Polypeptoids, or poly-N-substituted glycines, are a class of synthetic sequence-defined polymers that were initially developed as part of basic science efforts in peptidomimetics drug discovery.^{1,2} Peptoids are regioisomers of peptides with the side-chain attachment point shifted from the α -carbon to the amide nitrogen. This modification has enormous ramifications for the physicochemical properties of these molecules, due largely to the elimination of the C_α chiral center, removal of the $-NH$ hydrogen bond donor, and accessibility of both the cis and trans configurations of the ω dihedrals of the backbone amide bonds.^{3–5} This imparts peptoids with similar biocompatibility and chemical diversity as peptides but increased conformational flexibility, stability to proteolytic degradation, excellent thermal processability, and good backbone solubility.^{2,3,5–7} Moreover, peptoids can be synthesized cheaply and efficiently by submonomer solid-phase synthesis that enables precise control of monomer choice and placement.^{1,2,6} Peptoids belong to the class of “foldameric” molecules in the sense that they can adopt defined three-dimensional structures that can be controlled by the choice of the side chain.^{5,8,9} They also exhibit multi-molecular assembly into spheres, helices, and sheets, again controlled largely by the character of the side-chain groups and

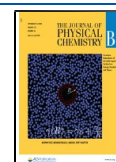
solvent environment.^{10–13} As such, peptoids constitute a class of highly tailorable molecular materials with broad applications in biomedical and materials science as peptidomimetic drugs, protein mimetic polymers, antimicrobials, surfactants, catalysts, and structure-directing agents.^{2,5–7,14–18}

Molecular simulation has played an important role in understanding and engineering peptoid behavior employing ab initio, all-atom, and coarse-grained (CG) simulations.⁵ Ab initio calculations have been used to accurately predict peptoid energy landscapes and stable conformations.^{19–21} All-atom force fields, typically parameterized against ab initio calculations and experimental data, have been developed and used to predict free-energy landscapes and stable conformations of single peptoid chains and peptoid crystals.^{9,19,22–24} All-atom models are too computationally expensive to simulate the self-assembly of multiple peptoid chains into larger aggregates, since doing so

Received: May 20, 2020

Revised: August 7, 2020

Published: August 10, 2020



can require consideration of hundreds of chains (several thousand atoms) over in excess of microsecond or even millisecond time scales.⁵ For this purpose, coarse-grained (CG) peptoid models have been developed that balance computational efficiency and chemical accuracy by lumping groups of atoms together into coarse-grained beads or pseudo-atoms.^{25–27} Doing so enables access to longer length and time scales by reducing the number of degrees of freedom that must be simulated, increasing the integration time step, and smoothing out some of the ruggedness of the underlying free-energy landscape that can accelerate the exploration of configurational space.²⁸

The first CG peptoid force field molecular foundry (MF)-CG-TOID was established by Zuckermann, Whitelam, and co-workers,²⁵ who were also responsible for one of the most popular all-atom peptoid models MFTOID.²³ MF-CG-TOID is an implicit solvent model that represents each peptoid monomer by two coarse-grained beads that possess not only a position but also an independently fluctuating orientational director. Incorporating anisotropy into the coarse-grained beads enables the use of a higher degree of coarse-graining but comes at the cost of inducing more complicated nonisotropic interaction potentials. The 115-parameter model was fitted in a bottom-up fashion against MFTOID all-atom simulation data and followed by top-down tuning against structural measurements of peptoid bilayers collected from experimental scattering measurements and all-atom molecular simulations. Importantly, the MFTOID all-atom model was itself carefully parameterized against *ab initio* calculations in which enhanced sampling calculations were employed to properly sample and model the slow *cis/trans* isomerization of the backbone ω dihedrals.²³ The MF-CG-TOID model was deployed to study the formation of bilayer sheets of $((\text{Nae-Npe})_{n/4}-(\text{Nce-Npe})_{n/4})$ “block-*n*” peptoids formed from alternating *N*-(2-aminoethyl)glycine, *N*-(2-phenylethyl)glycine, and *N*-(2-carboxyethyl)glycine monomers that have been the subject of robust experimental interest. The anisotropic nature of the interaction potentials is rather complicated and makes the model approximately 6 times more expensive than an isotropic analogue, but this cost is likely more than offset by the higher degree of coarse-graining that the anisotropy admits. The model is not transferable in the sense that its application to other peptoid side chains would require a complete reparameterization from scratch, and it is not designed to be compatible with existing coarse-grained force fields for solvents, proteins, lipids, and carbohydrates. As such, it is not readily possible to simulate conditions and systems beyond the specific peptoid–water system for which the model was parameterized.

Du, Rick, and Kumar developed a CG model of polysarcosine (poly(*N*-methyl glycine), *N,N*-dimethylacetamide) as a prototypical model for the polypeptoid backbone.²⁶ The optimized potential for liquid simulations-united atom (OPLS-UA) model is adopted for the intramolecular peptoid interactions that explicitly models all atoms except hydrogens, which are eliminated.²⁹ The water solvent is modeled using Molinero and co-worker’s model of water (mW) that models each water molecule as a single bead that interacts via a short-ranged Stillinger–Weber potential comprising two- and three-body terms.^{30,31} The primary parameterization effort within the paper was to fit the peptoid–peptoid and peptoid–water intermolecular potentials to all-atom reference simulations performed with the OPLS-AA force field.³² The degree of coarse-graining of the peptoid and water solvent is relatively mild, and the

solvent is explicitly represented, so it is largely the short-ranged nature of the intermolecular potential that enabled the observed 27-fold speedups. The CG model shows good agreement in predicting the density, surface tension, enthalpy of vaporization, isothermal compressibility, solvation free energy, and solvation structure of polysarcosine. This CG model provides a potential for the peptoid backbone and a foundation for the parameterization of derived models for other peptoid side chains. The model is designed to be compatible with the OPLS-UA force field and mW water model. Critically, however, the all-atom reference data for the bottom-up parameterization employed the OPLS-AA force field that is designed for peptides and proteins and not adapted to peptoids.³² Zuckermann, Whitelam, and co-workers developed their all-atom peptoid model MFTOID as a reparameterization of the CHARMM22 protein force field and showed that shifting the side-chain connection point from the α -carbon to the amide nitrogen necessitated substantial reparameterization of the bonded and nonbonded parameters of the amide core—partial charges, ω dihedral potential, Lennard-Jones interactions—to account for the much higher flexibility of the ω dihedral and absence of amide hydrogen bond donor.^{23,33,34} The OPLS-AA model for peptoids is not adapted in this way and its deficiencies are therefore propagated into the CG counterpart.

Most recently, Gao and Tartakovsky developed a CG model based on and compatible with the popular coarse-grained MARTINI model that is in wide use for the modeling of solvents, peptides, lipids, and carbohydrates.²⁷ Following MARTINI, approximately four heavy atoms are lumped into each coarse-grained bead.³⁵ The model parameters are fitted in a bottom-up fashion against all-atom simulation data employing the MFTOID all-atom peptoid model.²³ Bonded interactions are fitted against the all-atom distribution functions using direct Boltzmann inversion (DBI),³⁶ and nonbonded interactions are fitted to all-atom measurements of radii of gyration or solvation free energies. Four solvents are considered—water, acetonitrile, 1-octanol, and hexane—and modeled explicitly using existing MARTINI models. Although an appropriate all-atom peptoid model was employed for the bottom-up parameterization, no enhanced sampling of the backbone ω dihedrals over the course of the 2 μ s production runs was employed. Unlike peptides, the *cis* and *trans* conformations of the peptoid ω dihedral of the backbone amide bonds are—depending on the side-chain chemistry—approximately equally stable, and the ω dihedral readily transitions between these two torsional states.^{5,23,37} The activation energy for the transition, however, is on the order of 15 kcal/mol (25 $k_B T$ at $T = 300$ K) with this large barrier resulting in characteristic time scales for the *cis/trans* isomerization transition on the order of seconds or longer.^{5,38,39} As such, the ω dihedrals within the all-atom reference trajectories are very likely kinetically trapped within the initial isomerization state and the systematic errors associated with nonergodic sampling in these degrees of freedom are propagated into the CG model. These errors can result in substantially incorrect behaviors of the model. For example, this CG model predicts water to be poor solvent for polysarcosine (poly(*N*-methyl glycine)) chains that induces the chains to collapse into globules with a polymer scaling exponent of $\nu = 0.29 \pm 0.02$, whereas all-atom calculations predict water to be a good solvent and the chains to adopt swollen conformations with a scaling exponent of $\nu = 0.66 \pm 0.01$ consistent with experimental expectations.^{40–45} As we shall demonstrate, the CG model developed in this work predicts a scaling exponent of $\nu = 0.63 \pm 0.03$ in

excellent agreement with the all-atom calculations (Section 3.6.2).

In this work, we adopt a similar approach and motivation to Gao and Tartakovsky to develop a coarse-grained peptoid force field based on a bottom-up reparameterization of the MARTINI potential that is, by design, compatible with the universe of MARTINI models for solvents, peptides, lipids, and carbohydrates.²⁷ We fit the CG model parameters against all-atom simulations performed using the MFTOID potential that itself is an adaptation of the CHARMM22 potential modified for peptoids.^{23,33,34} We assure good sampling of the slow cis/trans isomerizations of the amide ω dihedrals in the all-atom simulations by performing enhanced sampling in all of these degrees of freedom using parallel bias metadynamics (PBMetaD).^{46,47} Bonded CG model parameters are fitted to the all-atom distribution functions using iterative Boltzmann inversion (IBI).^{48,49} IBI maintains the simplicity of DBI by treating each interaction potential independently; whereas, DBI inherently assumes the interaction potentials to be uncoupled, IBI implicitly treats their correlated dependencies by iterating to convergence.⁵⁰ Nonbonded CG model parameters are largely drawn from the standard MARTINI model with some additional tuning to fit the all-atom potentials of mean force (PMF) for peptoid chain extension as a form of PMF matching.⁵¹ We develop CG models for two representative peptoids: poly-sarcosine (poly(*N*-methyl glycine)) and poly(*N*-((4-bromophenyl)ethyl)glycine) (poly-Nbrpe). Polysarcosine is one of the earliest reported and simplest peptoids that serves as a prototypical model of a polypeptoid backbone.^{52–54} Poly-Nbrpe is part of a family of halophenyl side-chain peptoids that have been the subject of strong experimental interest as the hydrophobic block of amphiphilic diblock copolypeptoids that are capable of assembling into micron-sized sheets and tubes.^{55–58} Bromophenyl side chains in particular have drawn attention due to the capacity of bromine atom substituents to be visualized by cryogenic transmission electron microscopy.⁵⁸ The CG models of these two peptoids developed in this work are up to 25-fold more efficient than all-atom calculations, and we demonstrate their structural and thermodynamic predictions to be in excellent accord with all-atom calculations. This work establishes a new rigorously parameterized CG peptoid force field to enable the direct simulation of the behaviors and assembly of peptoid nanomaterials at length and time scales inaccessible to all-atom calculations.

2. THEORETICAL METHODS

In this section, we report (i) our simulation protocol for the all-atom molecular dynamics simulations, (ii) our implementation of parallel bias metadynamics to enhance sampling of the cis/trans isomerizations of the amide ω dihedrals in the all-atom simulations, and (iii) our coarse-grained molecular dynamics simulation protocol. The parameterization and validation of the coarse-grained model are described in Section 3.

2.1. All-Atom Molecular Dynamics Simulations. All-atom molecular dynamics simulations of polysarcosine (poly(*N*-methyl glycine)) and poly(*N*-((4-bromophenyl)ethyl)glycine) (poly-Nbrpe) peptoids were performed in Gromacs 2019.2.^{59,60} Peptoid molecules were capped with $-\text{CH}_3$ terminal groups, and initial configurations were constructed using Avogadro 1.2.0.⁶¹ Peptoids were modeled using the all-atom MFTOID force field developed by Zuckermann, Whitelam, and co-workers as a peptoid-tuned modification of the CHARMM22 peptide force field.^{23,33,34} In doing so, we make two assumptions.

First, the MFTOID peptoid backbone parameters are transferable between peptoids with different side chains. The force field was constructed with this modularity in mind, but it has not been conclusively demonstrated. Second, the peptoid side chains may be satisfactorily modeled using the same parameterization as in a peptide. A more rigorous approach would be to perform quantum mechanical calculations to explicitly parameterize the peptoid side-chain interaction potentials. We also note that during the performance of this work a newer all-atom peptoid model CGenFF that improves upon some of the property predictions of MFTOID was reported by Weiser and Santiso.²⁴ The CG parameterization protocol reported in this work could be used directly to develop a CG model fitted against CGenFF all-atom data. Water was modeled using the TIP3P model.⁶² Simulations of single (SC) sarcosine-5 and Nbrpe-5 peptoids were used to generate reference trajectories for the bottom-up parameterization of the CG model. We verified that the dependence of the bonded distribution functions upon chain lengths over the range $n = 5–15$ was very weak (see Section S1 of the Supporting Information). Peptoids were placed in a cubic box of dimensions $6 \times 6 \times 6 \text{ nm}^3$ and solvated with ~ 7000 water molecules to a density of 1.0 g/cm^3 . Additional single-chain simulations of sarcosine- n and Nbrpe- n , where $n = \{5, 7, 8, 10, 15, 25\}$, were performed in boxes from $6 \times 6 \times 6$ to $15 \times 15 \times 15 \text{ nm}^3$ to accommodate the longer chains and compute polymer scaling exponents. Simulations of multiple molecules used to simulate peptoid self-assembly were performed by placing 2–16 sarcosine-5 or Nbrpe-5 chains in a $4.5 \times 4.5 \times 4.5 \text{ nm}^3$ box and solvating with ~ 3000 water molecules to a density of 1.0 g/cm^3 . Periodic boundary conditions were applied in all three dimensions. Large atomic forces in excess of $1000 \text{ kJ}/(\text{mol} \cdot \text{nm})$ were removed through the steepest descent energy minimization. Initial atom velocities were assigned from a Maxwell–Boltzmann distribution at 300 K. The systems were then equilibrated for 200 ps in the NPT ensemble at 300 K and 1 bar, employing a velocity rescaling thermostat⁶³ with a time constant of 0.1 ps and a Berendsen barostat⁶⁴ with a time constant of 1.0 ps and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Production runs of 1 μs for the single-chain systems and 200 ns for the multichain systems were conducted in the NPT ensemble at 300 K and 1 bar, employing a Nosé–Hoover thermostat⁶⁵ with a time constant of 1.0 ps and Parrinello–Rahman barostat⁶⁶ with a time constant of 1.0 ps and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. A leap-frog algorithm⁶⁷ was used to integrate Newton's equation of motion with a 2 fs time step. Covalent bonds involving hydrogen atoms were fixed using the LINCS algorithm.⁶⁸ Long-range electrostatic interactions were treated by particle mesh Ewald summation⁶⁹ employing a real-space cutoff of 1.0 and 0.16 nm Fourier grid spacing that were subsequently optimized during runtime. Lennard-Jones interactions were smoothly shifted to zero at a cutoff of 1.2 nm. All-atom simulations used to generate reference data for CG model parameterization were subjected to enhanced sampling using parallel bias metadynamics, as described in Section 2.2. Simulation snapshots were saved for analysis at a period of 10 ps over the course of the production run and trajectories visualized in visual molecular dynamics (VMD).⁷⁰ Simulations were conducted on $5 \times 2.40 \text{ GHz}$ Intel Xeon Gold 6148 CPU cores and a 1 NVIDIA TITAN V GPU, achieving execution speeds for the single-chain 5-mer systems of $\sim 300 \text{ ns/day}$. All input files necessary to perform PBMetaD all-atom simulations of polysarcosine and poly-Nbrpe are provided in the Supporting Information.

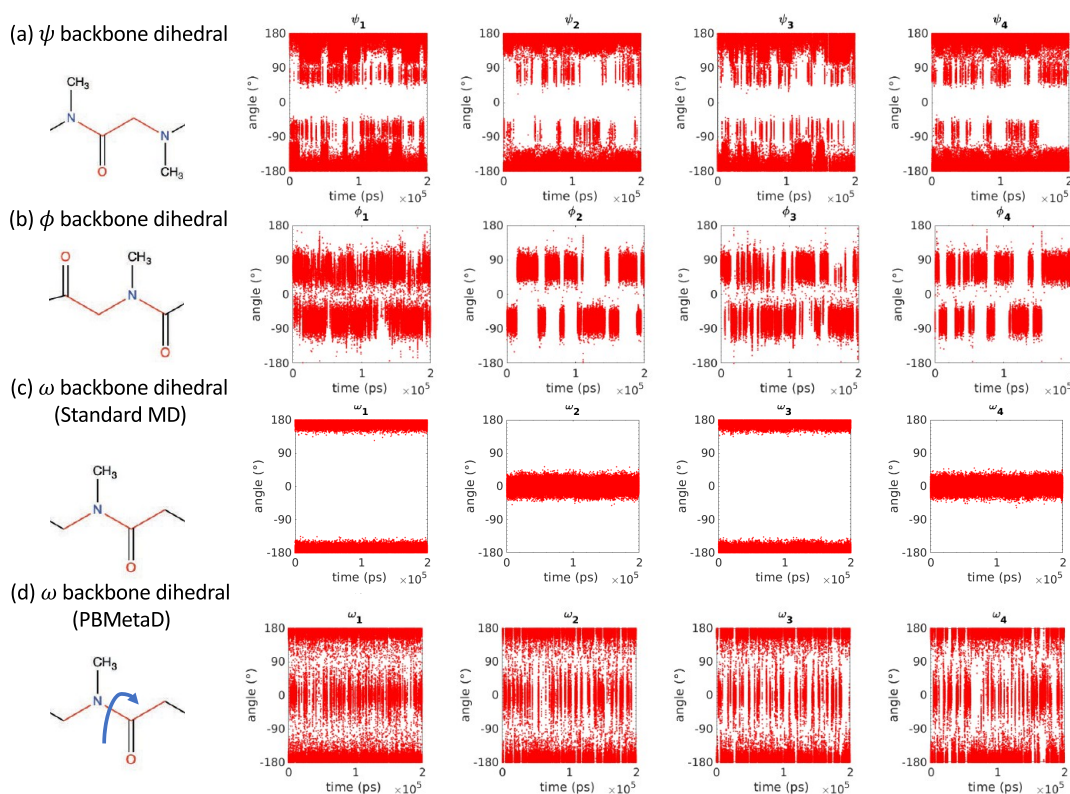


Figure 1. Time evolution of the backbone dihedral angles in 200 ns all-atom molecular dynamics simulations of sarcosine-5. In unbiased sampling, the (a) ψ and (b) ϕ dihedrals exhibit good exploration of their stable and metastable conformational states, whereas the (c) ω dihedrals are kinetically trapped in their initial cis or trans isomeric states. (d) Application of PBMetaD coupled to all four backbone ω dihedrals efficiently induces good sampling in these slow degrees of freedom.

2.2. All-Atom Parallel Bias Metadynamics. The amide ω dihedral in the peptoid backbone is far more flexible than in the peptide backbone due to the reduced double-bonded character of the peptoid tertiary amide compared to that of the peptide secondary amide.²¹ Unlike peptides, the ω dihedral generally samples both the cis and trans conformations, with the relative equilibrium proportions dictated by the nature of the peptoid side chain.⁵ The 15 kcal/mol ($25 k_B T$ at $T = 300$ K) activation barrier means that rotations around the ω dihedral are, unlike those around the ϕ and ψ backbone dihedrals, a rare event, with characteristic time scales of cis/trans isomerization exceeding seconds.^{5,38,39} As such, it is imperative to employ enhanced sampling techniques within all-atom molecular simulations to achieve ergodic sampling of the ω dihedrals and avoid kinetic trapping in the initial isomeric states.^{5,37} Figure 1a–c contrasts the good sampling of the ϕ and ψ backbone dihedrals with the kinetically trapped nature of the ω over the course of 200 ns unbiased molecular dynamics simulations of sarcosine-5.

Metadynamics is a powerful and popular enhanced sampling technique that works by depositing history-dependent Gaussian bias potentials in selected degrees of freedom to induce good sampling of rarely visited states.⁷¹ As is typical of collective variable acceleration techniques, the efficiency of metadynamics scales poorly with the number of slow degrees of freedom accelerated due to the exponential increase in the volume of phase space.⁴⁶ Parallel bias metadynamics (PBMetaD) is a variant of metadynamics that bypasses this deficiency by the simultaneous application of multiple one-dimensional (or low-dimensional) accelerating biases.^{46,47} Given a set of n collective variables $\vec{s}(\mathbf{r}^N) = \{s_i(\mathbf{r}^N)\}_{i=1}^n$, each a function of the configura-

tional state of the N particle system $\mathbf{r}^N \in \mathbb{R}^{3N}$, the time-dependent PBMetaD bias potential in $\vec{s}(\mathbf{r}^N)$ at time t is given by^{46,47,72}

$$U_{\text{PB}}(s(\mathbf{r}^N(t)), t) = -\frac{1}{\beta} \ln \sum_{i=1}^n \exp(-\beta U(\vec{s}(\mathbf{r}^N(t)), t)) \quad (1)$$

which comprises a summation over all one-dimensional time-dependent bias potential in each of the n collective variables $s_i(\mathbf{r}^N)$

$$U(s_i(\mathbf{r}^N(t)), t) = \int_0^t dt' W_i(t') \exp\left(-\frac{(s_i(\mathbf{r}^N(t)) - s_i(\mathbf{r}^N(t')))^2}{2\sigma_i^2}\right) W_{\text{PB}}(s_i(\mathbf{r}^N(t')), t') \quad (2)$$

and $\beta = 1/k_B T$ is the reciprocal temperature. The first two terms within the integral in eq 2 follow from the standard well-tempered metadynamics (WTMetaD),^{47,73} where

$$W_i(t') = W_0 \exp\left(-\frac{U(s_i(\mathbf{r}^N(t')), t')}{k_B \Delta T}\right) \quad (3)$$

is the time-dependent height of the deposited Gaussians in s_i , W_0 is the initial Gaussian height, σ_i is the time-independent width of the deposited Gaussians in s_i , $s_i(\mathbf{r}^N(t'))$ is the historical value of $s_i(\mathbf{r}^N)$ at time t' that defines the location of the deposited Gaussian, and the parameter ΔT controls how aggressively the

Gaussian heights are scaled down with time. The usual way to express ΔT is relative to the simulation temperature T through the bias factor $\gamma = (T + \Delta T)/T$. For $\gamma \rightarrow 1$, standard molecular dynamics is recovered (i.e., all deposited Gaussians have zero height), and for $\gamma \rightarrow \infty$, standard metadynamics is recovered (i.e., all deposited Gaussians have time-independent height W_0). The third term in eq 2

$$W_{\text{PB}}(\vec{s}(\mathbf{r}^N(t')), t') = \frac{\exp(-\beta U(\vec{s}(\mathbf{r}^N(t')), t'))}{\sum_{j=1}^n \exp(-\beta U(s_j(\mathbf{r}^N(t')), t'))} \quad (4)$$

is the conditional weight term that is central to PBMetaD. This term accounts for correlations among the n collective variables to conditionally weight the deposited Gaussian toward collective variables possessing lower instantaneous values of their one-dimensional bias potential and is critical in assuring correct convergence of the one-dimensional biasing potentials.⁴⁶ For finite $\gamma > 1$, the time-dependent bias becomes quasi-static as $t \rightarrow \infty$ and the simulation evolves according to a time-invariant bias potential.⁴⁶ Biased data collected within this regime may be straightforwardly reweighted back to the unbiased ensemble by applying the standard Torrie–Valleau umbrella sampling weight factor to each sampled configuration to negate the influence of the applied bias.^{46,74}

$$w(\vec{s}(\mathbf{r}^N)) \propto \exp(\beta U_{\text{PB}}(\vec{s}(\mathbf{r}^N))) \quad (5)$$

In this work, we apply PBMetaD to each all-atom peptoid simulation by coupling to all backbone ω dihedrals. In the case of poly-Nbrpe, we also coupled dihedral rotations around the central C–C bond of all ethyl side chains where we also observe poor sampling. We emphasize that the application of standard WTMD would fail for polypeptides longer than di- or tripeptides due to the poor scaling and difficulty in converging standard metadynamics when the number of accelerated degrees of freedom exceeds approximately 3. PBMetaD is vital in achieving proper sampling polypeptoid chains of any reasonable length and critical in supplying converged all-atom reference trajectories for our CG model parameterization. We implement PBMetaD in our simulations using the PLUMED plugin.⁷² Periodic bias potentials are applied to each ω dihedral over the range $x \in [-\pi, \pi)$. After some initial tuning, a Gaussian deposition pace of 500 ps, $W_0 = 1.2$ kJ/mol, $\sigma = 0.3$ rad, and $\gamma = 10.0$ were found to work well. The PBMetaD bias potentials were typically observed to converge and become quasi-static in approximately 10 ns of simulation as defined by the tandem criteria that the height of the deposited Gaussians plateaued to near-zero (i.e., $W_i(t) \rightarrow 0$, $\forall i$) and approximately uniform sampling in all accelerated dihedral angles was observed. Applying PBMetaD to sarcosine-5 produced excellent sampling of the four backbone ω dihedrals (Figure 1d), and reweighting of our data allowed us to estimate a free-energy barrier for the cis/trans isomerization of $\sim 20 k_B T$ at $T = 300$ K in good agreement with prior work.^{21,24,75}

2.3. Coarse-Grained Molecular Dynamics Simulations.

Coarse-grained molecular dynamics simulations of polysarcosine and poly-Nbrpe were conducted using Gromacs 2019.2.^{59,60} Peptoids were modeled using the CG force field developed in this work, as described in Section 3. All CG beads for the peptoids considered in the present work were electrically neutral, so the water solvent was modeled using the non-polarizable MARTINI water model.³⁵ Initial peptoid configurations were generated by PACKMOL.⁷⁶ Simulations of single

sarcosine- n and Nbrpe- n , where $n = \{5, 7, 8, 10, 15, 25\}$ chains, were conducted in cubic water boxes of dimension $6 \times 6 \times 6 \text{ nm}^3$ and multichain simulations of 2–16 sarcosine-5 and Nbrpe-5 in cubic boxes of dimension $4.5 \times 4.5 \times 4.5 \text{ nm}^3$. Water solvent was added to a density of 1.0 g/cm^3 . Periodic boundary conditions were applied in all three dimensions. Forces on beads in excess of $1000 \text{ kJ}/(\text{mol}\cdot\text{nm})$ were eliminated by the steepest descent energy minimization. Initial atom velocities were assigned from a Maxwell–Boltzmann distribution at 300 K. All simulations were conducted in the NPT ensemble at 300 K and 1 bar, employing a velocity rescaling thermostat⁶³ with a time constant of 1.0 ps and a Berendsen barostat⁶⁴ with a time constant of 3.0 ps and a compressibility of $3 \times 10^{-4} \text{ bar}^{-1}$. A leap-frog algorithm⁶⁷ was applied for integrating Newton's equation of motion with a 10 fs time step for polysarcosine and a 2 fs time step for poly-Nbrpe. A shorter time step was required to maintain the stability of the poly-Nbrpe system due to the presence of the small CC beads representing single CH_2 groups (cf. Section 3.1). Lennard-Jones interactions were smoothly shifted to zero at a cutoff of 1.2 nm. Since all CG beads were electrically neutral, these systems contained no Coulomb interactions. A series of single peptoid production runs of 200 ns were used to iteratively fit the CG model parameters. Multi-peptoid production runs of 1 μs were conducted to probe the peptoid self-assembly. Due to the soft nature of the CG intramolecular potentials that act over lumped groups of atoms, good sampling was observed in all CG degrees of freedom in unbiased molecular dynamics simulations without the need to implement any enhanced sampling. Simulation snapshots were saved for analysis at a period of 10 ps over the course of the production run and trajectories visualized in VMD.⁷⁰ Simulations were conducted on $5 \times 2.40 \text{ GHz}$ Intel Xeon Gold 6148 CPU cores and a 1 NVIDIA TITAN V GPU, achieving execution speeds for the single-chain 5-mer systems of $\sim 7.6 \mu\text{s}/\text{day}$ for polysarcosine and $\sim 1.5 \mu\text{s}/\text{day}$ for poly-Nbrpe. The corresponding execution speeds for the all-atom model reported in Section 2.1 of $\sim 300 \text{ ns}/\text{day}$ lead us to estimate a 25-fold speedup of the CG model for polysarcosine and a 5-fold speedup for poly-Nbrpe. As mentioned above, the presence of small CC beads representing single CH_2 groups mandated the use of a smaller integration time step to maintain numerical stability in the CG poly-Nbrpe model that diminishes its speedup relative to polysarcosine. All input files necessary to perform simulations of polysarcosine and poly-Nbrpe with our converged CG model are provided in the Supporting Information.

3. RESULTS AND DISCUSSION

In this section, we report our (i) choice of CG mapping from atoms to beads, (ii) parameterization of the CG model nonbonded interactions, (iii) parameterization of the CG model bonded interactions, and (iv) validation of the CG model against all-atom thermodynamic and structural benchmarks not used at any stage of the model parameterization: backbone potentials of mean force, polymer scaling exponents, dimerization potentials of mean force, PA–PA radial distribution functions (rdfs), and multichain self-assembly. A schematic flow diagram of the CG model parameterization and validation protocol is presented in Figure 2.

3.1. All-Atom to Coarse-Grained Mapping. The first step in our CG model parameterization is to define a lumping of atoms into CG beads (Figure 2, step 1). We present in Figure 3 our choice for this mapping for the polysarcosine and poly-Nbrpe models. The atoms lumped into each coarse-grained

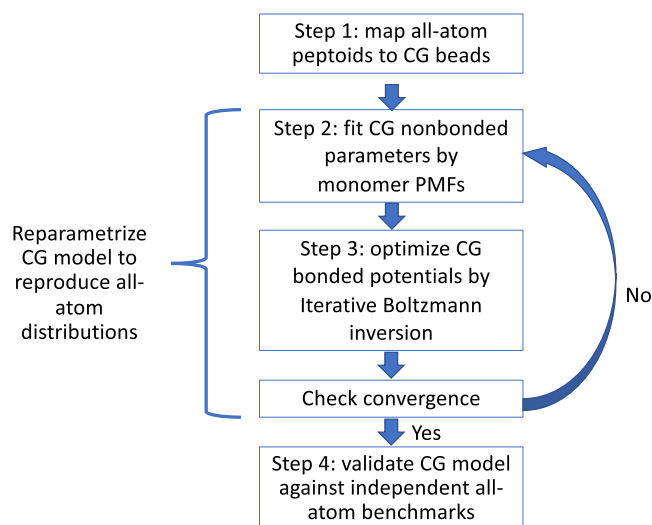


Figure 2. Flow chart of the bottom-up parameterization of CG peptoid force fields by iterative fitting of CG bonded and nonbonded distribution functions against all-atom reference data followed by independent validation.

bead under this mapping define the properties of the bead. For all of the CG beads in this work, we adopt the mass of the bead as the sum of the masses of the constituent atoms. If the interaction parameters of the bead do not exist within MARTINI, we fit the interaction parameters against all-atom simulation data. The mapping of the peptoid backbone follows that proposed by Gao and Tartakovsky designed for compatibility with the MARTINI model.^{27,35} Each CG backbone bead is denoted as PA and is centered on the midpoint of the C–N backbone bond. The chosen mapping for the PA bead coarse-grains over the cis/trans structural isomers, and we capture the changes in the interaction parameters of the bead in these two states by parameterizing the bonded and nonbonded PA interactions to reproduce the all-atom distribution functions averaged over the two structural isomers. We obtain the properly Boltzmann-weighted average over the cis/trans isomers by performing enhanced sampling of the ω dihedrals. The influence of the cis/trans isomerizations in

the CG model is therefore modeled implicitly through the fitted interaction parameters of the PA bead.

In the case of polysarcosine (Figure 3a–c), the PA bead encompasses the backbone N, C, and O atoms, half of each backbone CH₂ group, and the pendant –CH₃ side chain, giving the bead a mass of 71 amu that is close to the standard mass of 72 amu for MARTINI beads. For simplicity of the CG model, we make no distinction between PA beads within the core of the chain and those at the termini. As observed by Gao and Tartakovsky, introducing a new nonstandard MARTINI bead centered on the C–N midpoint produces a simpler distribution function, and therefore CG potential, for the PA–PA-bonded interaction.²⁷ In Figure 4, we show the backbone bond length

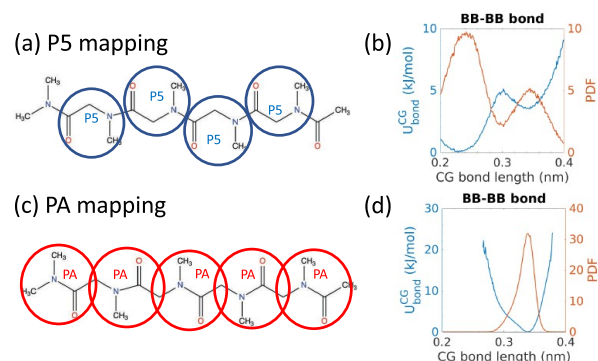


Figure 4. Comparison of the backbone bead bonded distribution functions and corresponding potential functions computed by the direct Boltzmann inversion extracted from all-atom simulations of sarcosine-5. A CG mapping of the backbone employing standard MARTINI P5 beads centered on the C_α (a) results in a bimodal distribution function for the bond length ((b) red curve) and corresponding double-well potential energy function ((b) blue). A CG mapping employing the new PA beads centered on the C–N bond midpoint (c) results in a unimodal distribution function ((d) red) and single-well, pseudo-harmonic potential ((d) blue).

distribution function extracted from sarcosine-5 all-atom molecular simulations and the corresponding potential function computed by DBI for a peptoid backbone represented by

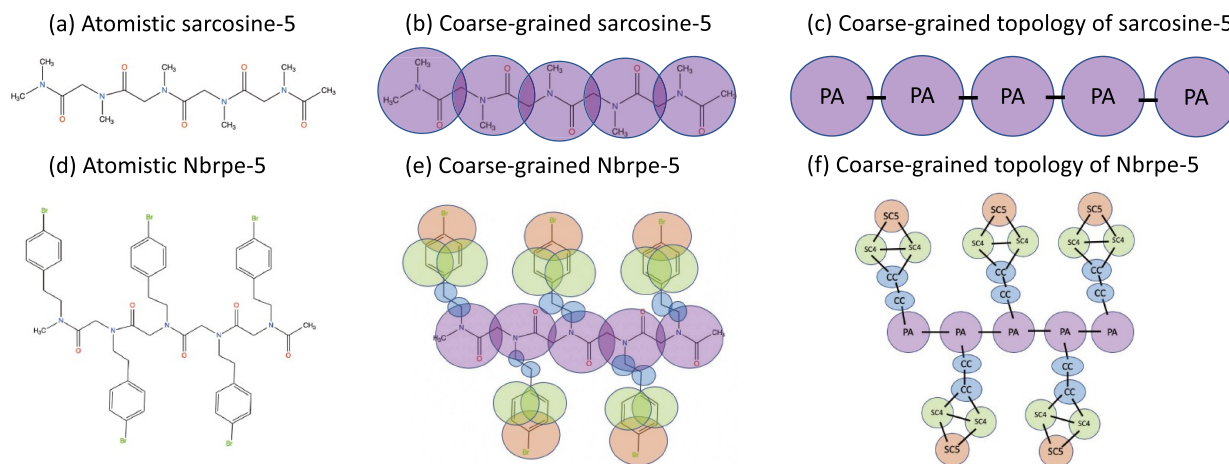


Figure 3. All-atom to coarse-grained mappings of polysarcosine and poly-Nbrpe. (a) All-atom model of the –CH₃-capped sarcosine-5. (b) All-atom to coarse-grained mapping of sarcosine-5. PA is a new nonstandard MARTINI bead whose bonded and nonbonded interactions are parameterized as part of this work. (c) CG representation of sarcosine-5. (d) All-atom model of the –CH₃-capped Nbrpe-5. (e) All-atom to coarse-grained mapping of Nbrpe-5. The SC4 and SC5 are standard MARTINI beads. CC is another nonstandard MARTINI bead that is parameterized in this work. (f) CG representation of Nbrpe-5. All chemical structures are generated in ChemDraw.⁷⁸

standard PS MARTINI beads centered on the C_α (Figure 4a,b)⁷⁷ compared to that using the new PA beads centered on the C–N bond midpoint (Figure 4c,d).²⁷ The primary advantage of the PA mapping over the PS is that it places the center of mass (COM) close to the geometric center of CG beads and admits a unimodal as opposed to bimodal distribution function and therefore a single-welled and pseudo-harmonic CG potential function. The cost, of course, is that the bonded and nonbonded interactions of the PA bead are not standard within MARTINI and must be fitted de novo from the all-atom reference simulations. The PA bead mapping and their interactions fully define the polysarcosine chain.

In the case of poly-Nbrpe (Figure 3d–f), we adopt a similar PA bead mapping of the backbone and a MARTINI-like coarse-graining of the side chain. The bromobenzene group is coarse-grained similar to the standard MARTINI model of chlorobenzene that includes one SC5 and two SC4 beads.⁷⁹ The bonded interactions of the SC4 and SC5 beads are, however, retuned from the standard MARTINI potentials using the IBI procedure detailed below. We experimented with a number of representations for the ethyl linker, which would be most naturally represented in MARTINI as a single C1 bead encompassing both CH_2 groups.^{35,77} We found, however, that it was critical to the success of the CG model in accurately representing the side-chain distribution functions that each CH_2 group was explicitly represented by its own nonstandard CC bead. The structural reason for this is straightforward—in the all-atom representation, the dihedral rotations around each of the three covalent bonds comprising the N–C–C–C ethyl linker allow the bromophenyl ring system to lie out of the plane of the peptoid backbone and also rotate relative to this plane. There is no way to represent this relative displacement without an explicit representation of the two CH_2 groups. Accurate representation of the relative positioning of the bromophenyl rings relative to the peptoid backbone is critical to accurate modeling of the interaction and self-assembly of Nbrpe peptoid chains and this motivates our representation of each CH_2 group as a CC bead. The nonbonded Lennard-Jones interactions of the CC bead are taken from the ethyl group in the CHARMM force field.⁸⁰ The penalty for our use of small, nearly atomistic, beads necessitated the use of a 2 fs integration time step to maintain numerical stability in the CG molecular simulation, which degrades the efficiency gains offered by the CG model relative to polysarcosine by approximately a factor of 5. We also note that the 14 amu mass of the CC beads is 5 times smaller than the standard 72 amu mass of MARTINI beads. Thermodynamic quantities do not depend on mass, but we may anticipate that this choice could lead to deviations in the dynamical time scales associated with this CG degree of freedom. The PA, CC, SC4, and SC5 bead mappings and their interactions fully define the poly-Nbrpe chain.

We choose to fit independent CG models for the polysarcosine and poly-Nbrpe systems using IBI to fit all bonded potentials and PMF matching to fit the nonbonded interactions. As such, the interaction potentials of the beads differ from their standard MARTINI parameterizations and the PA bead interaction potentials will differ between the two CG models. Viewing polysarcosine as the prototypical peptoid backbone, one could envision constructing a fully transferable CG peptoid model in which we employ the polysarcosine potential for the backbone and adopt side chain parameters from the standard MARTINI.²⁷ In the case of the poly-Nbrpe molecules studied in this work, we found this approach to be

insufficiently accurate, with the backbone parameters for poly-Nbrpe derived by bottom-up fitting against all-atom data differing from those for polysarcosine. This indicates, at least for the relatively bulky and short Nbrpe side chains, that the backbone–side-chain interaction may not admit a simple modular decomposition. Furthermore, our inability to adequately model the side-chain distribution functions in poly-Nbrpe under a standard MARTINI one-bead representation of the ethyl chain guards against a simple “drag-and-drop” of the MARTINI peptide representation without additional bottom-up reparameterization.

3.2. Coarse-Grained Model Parameterization. Following the definition of the mapping of atoms to CG beads (Figure 3), we proceed to fit the parameters of the CG model (Figure 2, steps 2 and 3). The CG interaction potential U^{CG} as a function of the positions of the N CG beads $\mathbf{r}^N$ is given by the sum of all constituent bonded and nonbonded interactions

$$U^{CG}(\mathbf{r}^N) = U_{\text{bonded}}^{CG}(\mathbf{r}^N) + U_{\text{non-bonded}}^{CG}(\mathbf{r}^N) \quad (6)$$

where

$$\begin{aligned} U_{\text{bonded}}^{CG}(\mathbf{r}^N) = & \sum_{i=1}^{n_{\text{bond}}} U_{\text{bond}}^{(i)}(l_i(\mathbf{r}^N)) + \sum_{j=1}^{n_{\text{angle}}} U_{\text{angle}}^{(j)}(\theta_j(\mathbf{r}^N)) \\ & + \sum_{k=1}^{n_{\text{dihedral}}} U_{\text{dihedral}}^{(k)}(\varphi_k(\mathbf{r}^N)) \\ & + \sum_{l=1}^{n_{\text{improper}}} U_{\text{improper}}^{(l)}(\zeta_l(\mathbf{r}^N)) \end{aligned} \quad (7)$$

$$U_{\text{non-bonded}}^{CG}(\mathbf{r}^N) = \sum_{p,q>p} U_{LJ}^{(p,q)}(|\mathbf{r}_p - \mathbf{r}_q|) \quad (8)$$

The bonded interactions comprise the sum of the potentials associated with the stretching of $i = 1, \dots, n_{\text{bond}}$ bond lengths l_i , bending of $j = 1, \dots, n_{\text{angle}}$ planar angles θ_j , rotations around $k = 1, \dots, n_{\text{dihedral}}$ backbone and side-chain dihedrals φ_k , and out-of-plane flexing of $l = 1, \dots, n_{\text{improper}}$ dihedrals ζ_l . Following MARTINI,³⁵ the nonbonded interactions comprise pairwise Lennard-Jones interactions that depend only on the linear distance $|\mathbf{r}_p - \mathbf{r}_q|$ between all pairs of CG beads $\{p, q\}$ that exist on different peptoid chains or separated by more than one bond on the same peptoid chain. (The only exception to this is the small CC bead in poly-Nbrpe for which nonbonded interactions are treated as described in Section 3.4.) None of the CG beads in our peptoid models carry any net charge, so we implement the nonpolarizable MARTINI water model and there are no electrostatic interactions in our model.³⁵ Extensions to systems comprising CG beads with net charge would require the incorporation of pairwise Coulomb interactions within the nonbonded interaction that are typically accounted for within MARTINI using reaction field electrostatics.⁸¹ It is also best practice for charged systems to employ the polarizable MARTINI water model.⁸²

We fit the CG model parameters against all-atom reference trajectories within an iterative procedure (Figure 2). All-atom simulations are performed using PBMetaD to achieve a converged sampling of the slow amide ω dihedrals and side-chain dihedral rotations, as described in Sections 2.1 and 2.2. CG simulations are performed after each update of the bonded and/or nonbonded parameters, as described in Section 2.3. In each pass through the loop, we first fit the free CG nonbonded interactions by matching the CG and all-atom PMF for the

extension of a single peptoid chain⁵¹ and then fit the free CG bonded interactions by matching the CG and all-atom bonded distribution functions using IBI.^{48,49} Iterating around this loop assures convergence of the interdependencies between the bonded and nonbonded interactions. The CG models for polysarcosine and poly-Nbrpe each converged in two passes through this loop.

3.2.1. Nonbonded Interactions: PMF Matching of Chain Extension. Following MARTINI, and in the absence of any net charge, the only nonbonded interactions between CG beads i and j are Lennard-Jones interactions (eq 8)³⁵

$$U_{LJ}^{(i,j)}(|\mathbf{r}_i - \mathbf{r}_j|) = U_{LJ}^{(i,j)}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (9)$$

requiring the specification of nonbonded pairwise interaction parameters ϵ_{ij} and σ_{ij} for each pair of bead types. The CG mappings of the two peptoids considered in this work are constructed from standard MARTINI beads SC4 and SC5 and nonstandard beads PA and CC³⁵ (Figure 3). The nonbonded interaction parameters for the SC4 and SC5 beads, together with the standard P4 beads of nonpolarizable MARTINI water, were adopted directly from the MARTINI force field.³⁵ The values of $\epsilon_{CC-CC} = 0.234$ kJ/mol and $\sigma_{CC-CC} = 0.36$ nm for the nonstandard CC bead representing a CH₂ group were adopted from the ethyl group in the CHARMM force field⁸⁰ and cross-interactions computed using Lorentz–Berthelot combining rules $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$ and $\sigma_{ij} = 0.5(\sigma_{ii} + \sigma_{jj})$.⁸³ This only leaves us to specify the PA bead nonbonded interaction parameters ϵ_{PA-PA} and σ_{PA-PA} , where again cross-interactions are defined by Lorentz–Berthelot. We fixed $\sigma_{PA-PA} = 0.47$ nm as is standard for the large MARTINI beads representing an approximately 4:1 heavy-atom-to-bead mapping.^{35,77} As such, ϵ_{PA-PA} is the only nonbonded interaction parameter that must be fitted.

During each pass of the iterative loop (Figure 2), we choose to fit ϵ_{PA-PA} by minimizing the least-squares difference between the PMF as a function of chain length ξ for the extension of a single 10-mer CG peptoid chain $W_{CG}(\xi)$ and that computed for an all-atom chain $W_{AA}(\xi)$

$$\epsilon_{PA-PA}^* = \operatorname{argmin}_{\epsilon_{PA-PA}} \int_{\xi_l}^{\xi_u} [W_{CG}(\xi; \epsilon_{PA-PA}) - W_{AA}(\xi)] d\xi \quad (10)$$

where $\xi_l = 1.1$ nm and $\xi_u = 2.5$ nm are the lower and upper limits of the chain length over which matching was performed, respectively, and $W_{CG}(\xi)$ and $W_{AA}(\xi)$ are the mean zeroed over the range $[\xi_l, \xi_u]$ to optimally align the profiles in a least-squares sense. Alternative fitting procedures such as matching the pairwise radial distribution between two solvated PA monomers could also have been adopted,^{84,85} but the matching procedure based on the single-chain PMF has the advantage of tuning the nonbonded parameter of the PA bead within a polymer chain—the environment in which the model will be deployed—as opposed to an isolated monomer. We have previously demonstrated that nonbonded parameters tuned in this manner for oligopeptide systems produce CG models in good agreement with all-atom calculations.^{86,87} In the CG simulations, ξ is simply defined as the center-of-mass distance between the terminal PA beads, whereas in the all-atom calculations, it is defined using the center-of-mass positions of the groups of atoms mapped to the terminal PA beads (Figure 3). We computed $W_{AA}(\xi)$ and $W_{CG}(\xi; \epsilon_{PA-PA})$ by performing umbrella sampling in ξ and

recovering estimates of the unbiased free-energy landscapes using the weighted histogram analysis method (WHAM).^{74,88,89} Umbrella windows were placed over the range $[\xi_l, \xi_u]$ in increments of $\Delta\xi = 0.2$ nm, and harmonic restraining potentials of 1000 kJ/(mol·nm²) were applied. Each umbrella window was equilibrated for 200 ps before conducting a 10 ns production run.

3.2.2. Bonded Interactions: Iterative Boltzmann Inversion. Initial estimates for each CG bonded interaction potential (eq 7) were generated by DBI of the corresponding distribution function extracted from the PBMetaD all-atom reference simulations under the defined CG mapping of atoms to beads (Figure 3)^{27,36,50,90}

$$U_{\text{bond}}^{(i)}(l_i) = -k_B T \ln \left(\frac{P_{\text{bond}}^{\text{AA}}(l_i)}{l_i^2} \right) + C_{l_i} \quad (11)$$

$$U_{\text{angle}}^{(j)}(\theta_j) = -k_B T \ln \left(\frac{P_{\text{angle}}^{\text{AA}}(\theta_j)}{\sin \theta_j} \right) + C_{\theta_j} \quad (12)$$

$$U_{\text{dihedral}}^{(k)}(\varphi_k) = -k_B T \ln(P_{\text{dihedral}}^{\text{AA}}(\varphi_k)) + C_{\varphi_k} \quad (13)$$

$$U_{\text{improper}}^{(l)}(\zeta_l) = -k_B T \ln(P_{\text{improper}}^{\text{AA}}(\zeta_l)) + C_{\zeta_l} \quad (14)$$

where $P_{\text{bond}}^{\text{AA}}(l_i)$, $P_{\text{angle}}^{\text{AA}}(\theta_j)$, $P_{\text{dihedral}}^{\text{AA}}(\varphi_k)$, and $P_{\text{improper}}^{\text{AA}}(\zeta_l)$ are the probability density functions for the bond length l_i , planar angle θ_j , proper dihedral φ_k , and improper dihedral ζ_l , respectively. The l_i^2 and $\sin \theta_j$ factors in the first and second expressions are the Jacobian determinants appropriate for the transformation from Cartesian coordinates,^{27,50} and the arbitrary additive constants C_{l_i} , C_{θ_j} , C_{φ_k} , and C_{ζ_l} are set such that the potential energy of the lowest energy state is zero. The superscripts on the CG interaction potential $U_{\text{bond}}^{(i)}$, $U_{\text{angle}}^{(j)}$, $U_{\text{dihedral}}^{(k)}$, and $U_{\text{improper}}^{(l)}$ indicate that in this work we fit a single potential for each type of interaction to keep the number of free parameters in the model under control. For example, a single potential $U_{\text{bond}}^{(PA-PA)}$ is fitted for all PA–PA bond stretches in the peptoid backbone regardless of whether they were located in the middle or end of the chain by averaging over all such probability distribution functions. We found this environment-independent approach to work well for the peptoids considered in this work.

The CG potentials computed by DBI will exactly reproduce the all-atom distribution functions only if there are no dependencies between these degrees of freedom and each distribution function contains no contaminating influences due to other interaction potentials.⁵⁰ In general, this is not the case and we progressively refine our estimates of the CG potentials to eliminate these dependencies using IBI.^{48,49} During each pass of the iterative loop (Figure 2), we apply the IBI update equation to each CG bonded interaction $U(x)$

$$U^{(t+1)}(x) = U^{(t)}(x) + \alpha k_B T \ln \left(\frac{P^{\text{CG}}(x; U^{(t)}(x))}{P^{\text{AA}}(x)} \right) \quad (15)$$

where $U^{(t)}(x)$ is the current CG potential, $U^{(t+1)}(x)$ is the updated potential, $P^{\text{AA}}(x)$ is the corresponding probability density function extracted from the PBMetaD all-atom reference trajectory under the CG mapping, $P^{\text{CG}}(x; U^{(t)}(x))$ is the analogous probability density function computed from a CG simulation conducted with CG potential $U^{(t)}(x)$, and $\alpha \in (0, 1]$ is a scaling factor to stabilize the update that we set to unity.⁹¹ We perform new CG molecular simulations during each pass

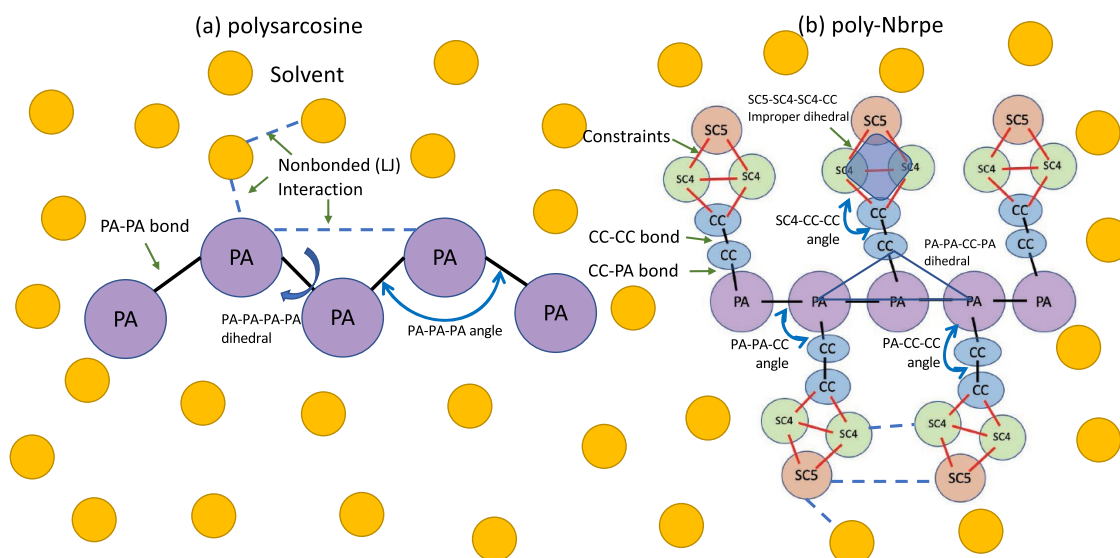


Figure 5. Schematic illustration of the bonded and nonbonded interactions in CG models of (a) polysarcosine and (b) poly-Nbrpe in the solvent. For visual clarity, the PA–PA bond, PA–PA–PA angle, and PA–PA–PA–PA dihedral interactions are not explicitly represented in poly-Nbrpe. The CG solvent beads are colored yellow.

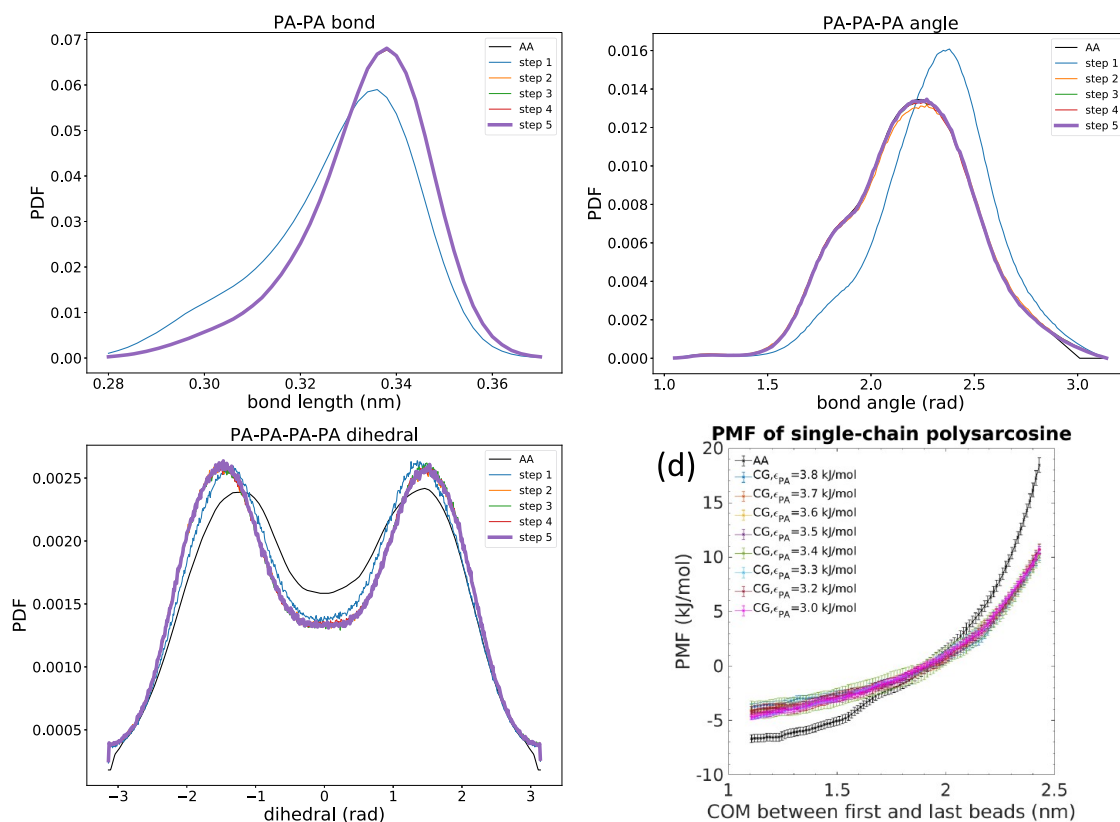


Figure 6. Convergence of the CG model for polysarcosine. (a–c) Convergence of the PA–PA bond, PA–PA–PA angle, and PA–PA–PA–PA dihedral distribution functions to the all-atom targets. (d) Nonbonded PA–PA Lennard-Jones parameter stabilizes to $\epsilon_{\text{PA-PA}} = 3.3$ kJ/mol corresponding to a weak optimum in the final round of model fitting that produces reasonable agreement of the CG PMF to the all-atom result. The PMF profiles are plotted with zero mean corresponding to the optimal mutual alignment in a least-squares sense.

using the protocol described in Section 2.3 and automate the IBI update procedure using the coarse-graining toolkit of the Versatile Object-oriented Toolkit for Coarse-graining Applications software package (VOTCA-CSG).^{91,92} For representational flexibility, we represent each $U_{\text{bond}}(l)$ and $U_{\text{angle}}(\theta)$ as tabulated potentials. The domain of the tabulated potentials for

the angles is $[0, \pi)$. The domain of the bond length potentials is defined implicitly through the domain of the target distribution and linearly extrapolated up to 3 nm by default. Any bond length exiting the parameterized domain during runtime will cause the simulation to stop. The periodicity of the proper dihedrals makes them more stably represented by analytical functions, and

we find that the Ryckaert–Bellemans potential comprising a six-term expansion in powers of $\cos \varphi$ provides an accurate and flexible representation^{27,93}

$$U_{\text{dihedral}}(\varphi) = \sum_{n=0}^5 (-1)^n C_n (\cos \varphi)^n \quad (16)$$

Interestingly, we found that the calculated potentials for the proper and improper dihedrals changed very little from the DBI result in successive IBI passes, suggesting that these distribution functions are approximately decoupled from the other interactions for both polysarcosine and poly-Nbrpe. The converged $U_{\text{dihedral}}(\varphi)$ and $U_{\text{improper}}(\zeta)$ are therefore largely unchanged from the DBI estimates, whereas the other interactions were substantially refined by IBI. Our system contains only a single improper interaction governing the SC5–SC4–SC4–CC ring system within the Nbrpe side chain, and for simplicity in our fitted force field, we treated this improper interaction with the corresponding MARTINI parameters for a tryptophan side chain without any significant loss of accuracy.⁹⁴

3.3. Coarse-Grained Model for Polysarcosine. A CG model of polysarcosine was generated according to the protocol detailed in Section 3.2 (Figure 2). The simplicity of the CG mapping of polysarcosine means that we fit only three CG bonded interactions—PA–PA bond, PA–PA–PA angle, and PA–PA–PA–PA dihedral—in addition to the nonbonded Lennard-Jones epsilon parameter $\epsilon_{\text{PA-PA}}$ (Figure 5a). All interactions converge after two passes through the iterative update loop. Convergence of the bonded and nonbonded distribution functions through the final pass is illustrated in Figure 6 and shows good agreement with the all-atom target distributions. Convergence is defined as changes in successive updates of less than 0.5 kJ/mol in the nonbonded interaction energy and 0.3 kJ/mol at any point in the bonded potential functions. The nonbonded interaction converges to $\epsilon_{\text{PA-PA}} = 3.3$ kJ/mol, and the expansion coefficients for the converged Ryckaert–Bellemans potential for $U_{\text{dihedral}}^{(\text{PA-PA-PA-PA})}$ are reported in Table 1. The complete potential is provided within the CG force field files in the Supporting Information.

Table 1. Coefficients of the CG Ryckaert–Bellemans Potential for $U_{\text{dihedral}}(\varphi)$ in Polysarcosine (Equation 16)^a

proper dihedral	C_0	C_1	C_2	C_3	C_4	C_5
PA–PA–PA–PA	0.3175	1.0597	1.3651	−0.8711	1.6516	1.8755

^aValues are reported in kJ/mol.

3.4. Coarse-Grained Model for Poly-Nbrpe. A CG model of poly-Nbrpe was also developed in an analogous manner to polysarcosine but where the (4-bromophenyl)ethyl side chains require parameterization of additional interactions (Figure 5b). In terms of bonded interactions, the CG mapping comprises interactions within the backbone—PA–PA bond, PA–PA–PA angle, and PA–PA–PA–PA dihedral—within the side chain—CC–CC bond, SC4–CC–CC angle, and SC5–SC4–SC4–CC improper dihedral—and between the backbone and side chain—CC–PA bond, CC–CC–PA angle, CC–PA–PA angle, and PA–PA–CC–PA dihedral. The SC5–SC4, SC4–SC4, and SC4–CC bond lengths within the side-chain ring system are fixed, as is standard practice for ring systems within MARTINI,⁷⁷ and the SC5–SC4–SC4–CC improper dihedral

is adopted from the MARTINI parameterization of tryptophan.⁹⁴ The small CC bead representing a CH₂ group is essentially a united atom, and so for these beads, we also account for 1–4 nonbonded interactions within the peptoid chain to prevent aphysical overlaps.²⁷ The only nonbonded parameter to be fitted is $\epsilon_{\text{PA-PA}}$.

We performed an asynchronous update of the bond stretching potentials and planar angle and dihedral potentials to avoid the numerical instability we observed in fitting them simultaneously within the IBI update. Specifically, we first converge the bond stretching potentials and then converge the angles and dihedrals. We illustrate in Figure 7 the convergence of the bonded and nonbonded distribution functions within the second and final pass through the iterative loop in Figure 2. The bond stretching potentials converged in 10 IBI iterations and the angle and dihedral potentials in six iterations. We then tested whether the bond stretching distributions were sufficiently perturbed by the angle and dihedral updates to require IBI refitting. The CC–CC and CC–PA distributions were essentially unchanged, and the sharply peaked PA–PA distribution underwent only a +0.003 nm shift, which we deemed sufficiently small that this was unnecessary. We observe only relatively minor systematic deviations from the all-atom distribution functions in the PA–PA–PA–PA dihedral, CC–PA–PA angle, and PA–PA–CC–PA dihedral introduced by the coarse-grained mapping. The most substantive discrepancy is the inability of the CG model to reproduce the bimodality at each end of the range of the PA–PA–CC–PA dihedral. It is not clear to us precisely which feature of the CG mapping is responsible for this deficiency. The optimum value of $\epsilon_{\text{PA-PA}} = 3.0$ kJ/mol is in good agreement with the corresponding value of $\epsilon_{\text{PA-PA}} = 3.3$ kJ/mol for polysarcosine. The Ryckaert–Bellemans expansion coefficients for $U_{\text{dihedral}}^{(\text{PA-PA-PA-PA})}$ and $U_{\text{dihedral}}^{(\text{PA-PA-PA-PA})}$ are reported in Table 2. The complete potential is provided within the CG force field files in the Supporting Information.

3.5. Time Scale of Coarse-Grained Simulations. The CG model offers acceleration over all-atom calculations by lumping atoms into beads, which not only reduces the number of degrees of freedom to be simulated but also smooths out the underlying free energy landscape leading to artificially accelerated motion through configurational space.²⁸ Different degrees of freedom may be affected to different degrees, and both speedups and slow-downs of CG time scales relative to all-atom have been observed.²⁸ We quantify the translational acceleration as the most important motion governing molecular collisions and the kinetics of self-assembly through the ratio of the coarse-grained D_{CG} and all-atom D_{AA} self-diffusion coefficients.^{51,95–97} Self-diffusion constants were estimated by computing the mean-squared displacement (MSD) of the center of mass of single 5-mer peptoid chains in unbiased molecular simulations in water solvent at 300 K and 1 bar (Sections 2.1 and 2.3) and applying the Einstein relation to the linear regime.⁹⁸ The calculated speedups for sarcosine-5 and Nbrpe-5 are (1.09 ± 0.34) and (0.96 ± 0.46) , respectively (Table 3), indicating that the dynamical time scales of the all-atom and coarse-grained runs have no significant difference. Consequently, we report all of the dynamical quantities calculated in the coarse-grained runs based on the unscaled simulation time.

3.6. Validation of the Coarse-Grained Models. Having fitted CG models for polysarcosine and poly-Nbrpe, we now proceed to validate their predictions against five all-atom benchmarks not used during any stage of model fitting (Figure 2, step 4): (i) backbone PMFs, (ii) polymer scaling exponents, (iii)

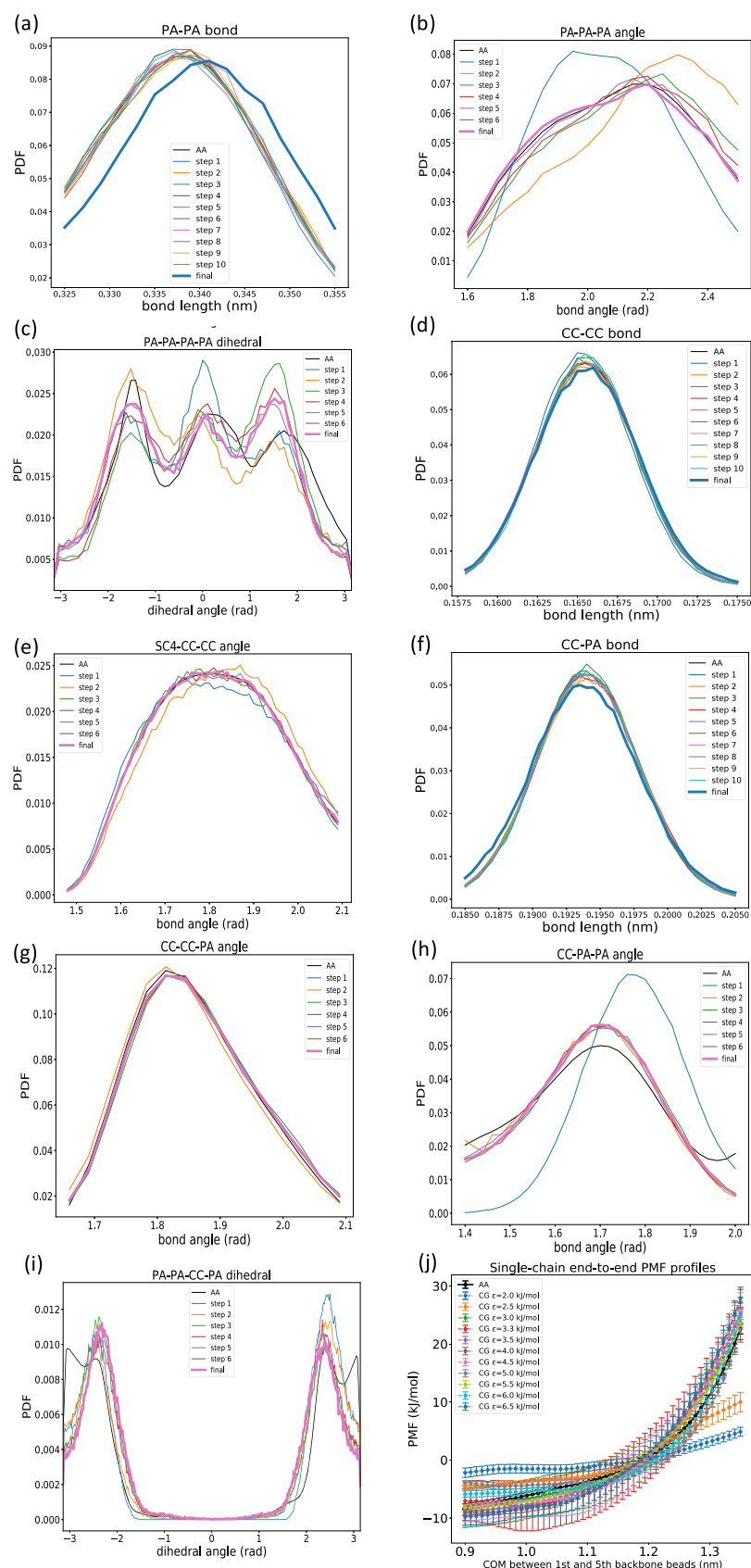


Figure 7. Convergence of the CG model for poly-Nbrpe. Convergence of the (a–c) backbone PA–PA bond, PA–PA–PA angle, and PA–PA–PA dihedral, (d, e) side-chain CC–CC bond and SC4–CC–CC angle, and (f–i) backbone–side-chain CC–PA bond, CC–CC–PA angle, CC–PA–PA angle, and PA–PA–CC–PA dihedral distribution functions to the all-atom targets. (j) Nonbonded PA–PA Lennard-Jones parameter stabilizes to $\epsilon_{\text{PA-PA}} = 3.0$ kJ/mol that produces very good agreement of the CG PMF to the all-atom result within the final round of fitting. The PMF profiles are plotted with zero mean corresponding to the optimal mutual alignment in a least-squares sense.

Table 2. Coefficients of the CG Ryckaert–Bellemans Potential for $U_{\text{dihedral}}(\varphi)$ in Poly-Nbrpe (Equation 16)^a

proper dihedral	C_0	C_1	C_2	C_3	C_4	C_5
PA–PA–PA–PA	0.5709	0.0622	3.0820	0.3375	−1.7662	1.0332
PA–PA–CC–PA	6.8729	−7.5264	−5.3617	−0.2755	5.1512	1.8120

^aValues are reported in kJ/mol.**Table 3.** Speedup of the Coarse-Grained Model Computed from the Ratio of the Coarse-Grained D_{CG} and All-Atom D_{AA} Self-Diffusion Coefficients^a

molecule	$D_{\text{AA}} \times 10^6$ (cm ² /s)	$D_{\text{CG}} \times 10^6$ (cm ² /s)	speedup ($D_{\text{CG}}/D_{\text{AA}}$)
sarcosine-5	10.17 ± 1.88	11.05 ± 1.45	1.09 ± 0.34
Nbrpe-5	2.61 ± 0.59	2.51 ± 0.63	0.96 ± 0.46

^aUncertainties in self-diffusion coefficients correspond to 95% confidence intervals in the gradient of the least-squares fit of the Einstein relation to the linear region of the MSD curve. Uncertainties in the speedup are computed by the propagation of errors.

PMFs of dimerization, (iv) backbone radial distribution functions, and (v) kinetics and morphology of poly-Nbrpe self-assembly. We select for the first validation a sufficiently small system that we can perform converged enhanced sampling in all backbone ω and side-chain C–C dihedrals in the all-atom calculations. In the remaining four validations, the all-atom systems contain sufficiently long peptoid chains and/or sufficiently many chains—some containing up to 160 such dihedrals—that it is computationally intractable to achieve converged enhanced sampling by simultaneously accelerating so many slow degrees of freedom. In these validations, we make the pragmatic choice to probe the agreement of the models in such large systems by conducting unbiased all-atom calculations. As we shall see, we observe good agreement between the all-atom and CG models, but this must be viewed under the proviso that the all-atom dihedrals are kinetically trapped over the course of these short runs. This deficiency is mitigated by, and the good agreement likely due to, two factors. First, the coarse-grained mapping of the all-atom data (Figure 3) integrates over multiple all-atom degrees of freedom and so attenuates the influence of any one kinetically trapped degree of freedom in the calculated CG observables. Second, we initialize the all-atom runs with a random initial distribution of dihedral angles, so that averaging over long peptoid chains or multiple peptoid chains averages over multiple quenched, although not necessarily Boltzmann-distributed, realizations of the dihedral states.

3.6.1. Backbone PMFs of Sarcosine-5 and Sarcosine-7.

First, we compare the conformational free-energy landscapes sampled by the all-atom and CG models of sarcosine-5 and sarcosine-7 constructed in two PA–PA–PA–PA dihedral angles. We conduct all-atom simulations in water at $T = 300$ K and $P = 1$ bar using PBMetaD to accelerate sampling of all backbone ω and side-chain C–C dihedrals, as described in Section 2.2. We reweight the biased simulation data under the all-atom to CG mapping (Figure 3) to compile two-dimensional (2D) histograms $P^{\text{AA}}(\varphi_1, \varphi_2)$ in the two PA–PA–PA–PA dihedral angles $\{\varphi_1, \varphi_2\}$ in sarcosine-5 and in the two central $\{\varphi_2, \varphi_3\}$ and two distal $\{\varphi_1, \varphi_4\}$ of the four such PA–PA–PA–PA dihedral angles $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$ in sarcosine-7. We then construct the 2D all-atom PMF using the statistical mechanical identity $\beta F^{\text{AA}} = -\ln P^{\text{AA}} + C$, where $\beta = 1/k_{\text{B}}T$ is the reciprocal temperature, F is the Gibbs free energy, and C is an arbitrary additive constant. We conduct 1 μs unbiased CG simulations to

achieve good conformational sampling in all degrees of freedom and construct the CG PMFs βF^{CG} .

The comparison of the 2D PMFs for sarcosine-5 in Figure 8a,b shows good agreement within $\pm 1 k_{\text{B}}T$ in the depths of the free-energy wells between the all-atom and CG models, but that the location of the wells in the (φ_1, φ_2) -plane in the all-atom model is not captured by the CG model that predicts a landscape with a simple fourfold rotational symmetry with local minima at $(\pm\pi/2, \pm\pi/2)$. This discrepancy is attributable to our fitting of a single position-independent potential for each CG degree of freedom, whereas in the all-atom model, there exist position-dependent correlations that are manifested here in the interaction between two adjacent PA–PA–PA–PA dihedral angles. This effect is clearly illuminated in the analysis of sarcosine-7, where the PMFs in the two central PA–PA–PA–PA dihedrals (Figure 8c,d) show similar trends to sarcosine-5 due to the correlated interactions in the all-atom model. In contrast, the PMFs in the two distal PA–PA–PA–PA dihedrals (Figure 8e,f) show excellent agreement between the all-atom and CG models due to the approximately independent behavior of these two angles in the all-atom and CG models. This analysis demonstrates that the CG and all-atom backbone conformational free-energy landscapes for short polysarcosine chains are in excellent agreement in terms of the location and depth of the local minima for dihedral angles that are sufficiently distant to behave independently. Correlated couplings between adjacent dihedral angles in the all-atom model cannot be captured by the position-independent potentials in the fitted CG model and, although the depth of the local minima is well represented, it does not reproduce the full richness of the all-atom landscape. This disagreement is an intrinsic deficiency of a CG model employing position-independent potential functions.

3.6.2. Polymer Scaling Exponents. Second, we compare the structure of isolated peptoid chains in water. Flory theory predicts how the size of a polymer R scales with the degree of polymerization N as a function of solvent quality through the scaling relation $R \sim N^\nu$.⁴⁰ Water is known to be a good solvent for polysarcosine with $\nu \approx 3/5$ and a poor solvent for poly-Nbrpe with $\nu \approx 1/3$, presenting a means to test the predictions of our CG model.^{27,40,55,57,58} We present in Figure 9 log–log plots of the dependence of the radius of gyration R_{g} of polysarcosine and poly-Nbrpe chains with degrees of polymerization spanning $N = 5, 7, 8, 10, 15, 25$ computed using the all-atom and CG models at $T = 300$ K and $P = 1$ bar. We fit scaling exponents over this range and observe very good agreement in the scaling exponents between the all-atom and CG models and reasonable correspondence with Flory theory, with $\nu_{\text{sarcosine}}^{\text{AA}} = 0.66 \pm 0.01$ and $\nu_{\text{sarcosine}}^{\text{CG}} = 0.63 \pm 0.03$, and $\nu_{\text{Nbrpe}}^{\text{AA}} = 0.24 \pm 0.01$ and $\nu_{\text{Nbrpe}}^{\text{CG}} = 0.27 \pm 0.01$. It is computationally expensive to collect all-atom data at large N but calculations using the CG model at $N = 40, 70, 100$ show that the extrapolated scaling trends are preserved to large N . We do observe the larger uncertainties in the large- N polysarcosine calculations where convergence of the structural averages is relatively slow due to the large configurational space available to the long swollen chains. The discrepancy of the computed scaling exponents with

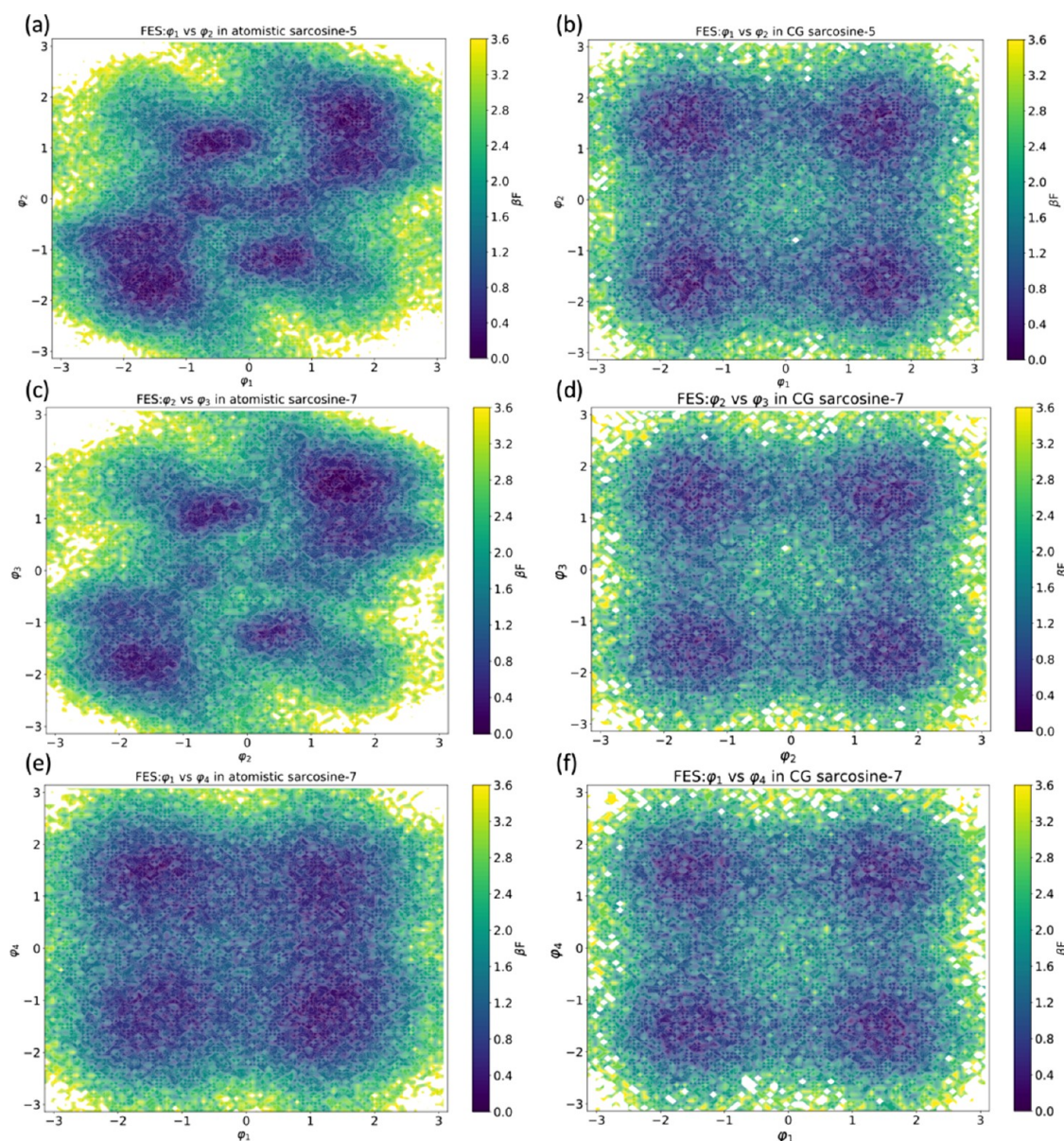


Figure 8. PMFs in pairs of PA–PA–PA–PA dihedral angles for sarcosine-5 and sarcosine-7 in water at $T = 300$ K and $P = 1$ bar. The 2D PMFs in the two sarcosine-5 PA–PA–PA–PA dihedral angles $\beta F(\varphi_1, \varphi_2)$ for (a) all-atom and (b) CG sarcosine-5. The 2D PMFs in the two central PA–PA–PA–PA dihedral angles $\beta F(\varphi_2, \varphi_3)$ for (c) all-atom and (d) CG sarcosine-7, and in the two distal PA–PA–PA–PA dihedral angles $\beta F(\varphi_1, \varphi_4)$ for (e) all-atom and (f) CG sarcosine-7. The arbitrary zero of free energy in each plot is fixed at the global free-energy minimum of the surface.

Flory theory is attributable to the nonideal nature of the chains, which contain complex bonded and nonbonded interactions.⁴⁰ This provides validation that our CG model is correctly predicting the structure of single polypeptoid chains in water solvent. We note that although the slopes of the all-atom and CG curves in Figure 9 are in good agreement, there is a systematic offset of the CG model to lower values of R_g . This can be easily understood as the result of the CG mapping that collapses atoms to the center of CG beads and therefore effectively reducing the spatial extent and pervaded volume of the chain.

To provide additional structural comparisons between the all-atom and CG chains, we also report in Figure 9 the principal moments of the gyration tensor $\{\lambda_1, \lambda_2, \lambda_3\}$ that measure the linear extent of the chain along its three principal axes and are related to the radius of gyration as $R_g^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$.⁹⁹ We observe good agreement within error bars between all principal moments computed from the all-atom and CG models, with the

exception of the shortest ($N \leq 10$) polysarcosine chains. For these chains, the leading moment λ_1 characterizing the longest linear extent of the chain is in good agreement, but the CG model systematically underpredicts λ_2 and λ_3 . The origin of this trend can also be attributed to the CG mapping that collapses atoms to the center of the CG beads, the effect of which is most pronounced as a large relative effect in the very small values of λ_2 and λ_3 for the shortest polysarcosine chains.

3.6.3. PMF of Dimerization. Third, we test thermodynamic predictions of the PMF of dimerization as a measure of the free-energy change in bringing together two peptoid chains. It is known that polysarcosine chains tend to remain solvated in water due to their hydrophilic nature and propensity to swell in water as a good solvent, whereas poly-Nbrpe chains tend to aggregate due to their hydrophobic collapse in water as a poor solvent.^{55,57,58} We therefore anticipate that the dimerization PMF should be positive (unfavorable) for polysarcosine and

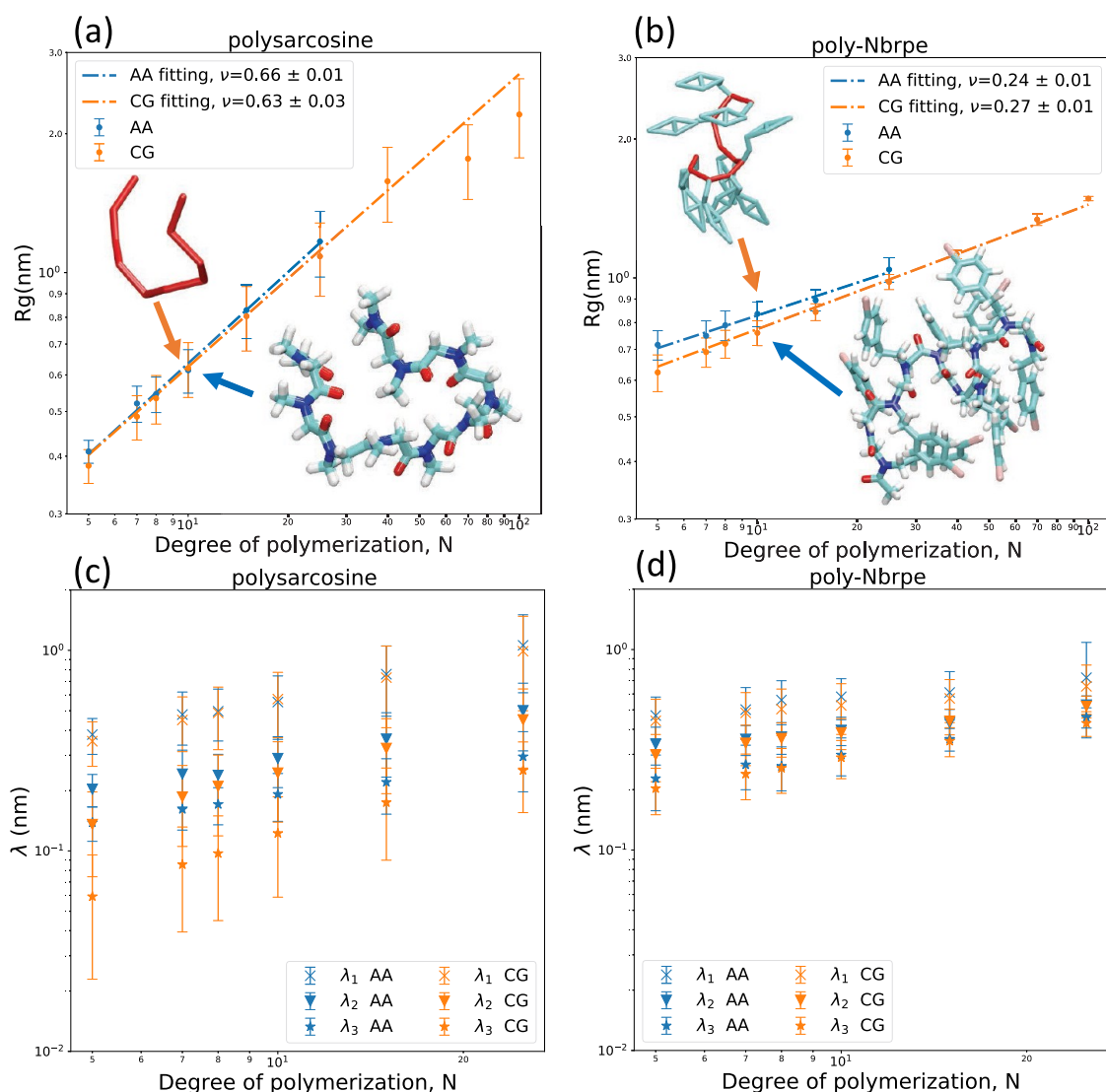


Figure 9. Polymer scaling exponents of polysarcosine and poly-Nbrpe in water at $T = 300$ K and $P = 1$ bar. Radius of gyration R_g of (a) polysarcosine and (b) poly-Nbrpe in water as a function of chain length computed in all-atom and CG molecular simulations. The scaling exponents of $\nu_{\text{sarcosine}}^{\text{AA}} = 0.66 \pm 0.01$ and $\nu_{\text{sarcosine}}^{\text{CG}} = 0.63 \pm 0.03$ are in good mutual agreement and consistent with water as a good solvent for these chains. The scaling exponents of $\nu_{\text{Nbrpe}}^{\text{AA}} = 0.24 \pm 0.01$ and $\nu_{\text{Nbrpe}}^{\text{CG}} = 0.27 \pm 0.01$ are also in good mutual agreement and consistent with water as a poor solvent. Standard errors in the scaling exponents are estimated by the standard deviation over leave-one-out fits. Principal moments of the gyration tensor $\{\lambda_1, \lambda_2, \lambda_3\}$ of (c) polysarcosine and (d) poly-Nbrpe measuring the linear extent of the chain along its three principal axes and related to the radius of gyration as $R_g^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$. Error bars on each point indicate standard deviations over the course of 1 μs simulations. Representative snapshots of 10-mer all-atom and CG peptides are shown.

negative (favorable) for poly-Nbrpe. We compute dimerization PMFs for pairs of sarcosine-5 and Nbrpe-5 chains in water at $T = 300$ K and $P = 1$ bar by performing umbrella sampling in the center-of-mass (COM) separation d between the central (i.e., third) backbone PA bead, defined in the all-atom model by applying our CG mapping (Figure 3).^{51,74,87} Umbrella sampling was conducted over the range of $d = 0.5$ – 2.7 nm in increments of 0.05 nm. Harmonic biasing potentials with force constants of 5000 kJ/(mol·nm²) were applied in each window over the course of 10 ns production runs. The unbiased PMFs were estimated from the biased trajectories using the weighted histogram analysis method (WHAM)^{88,100} implemented in the *g_wham*⁸⁹ module of GROMACS 2019.1 and incorporating the entropic correction.^{59,60}

The PMF curves reported in Figure 10 show good agreement between the all-atom and CG models in terms of the large-scale

features of the curves and the predicted free-energy changes, and good correspondence in the position of the minimum at $d = 1.0$ nm for the Nbrpe-5 chains. Discrepancies on the smaller scale (e.g., the shallow local minimum at $d = 1$ nm in the CG model of sarcosine-5 that is absent in the all-atom) may be attributed to the inherent information loss and approximations associated with the coarse-grained mapping. Defining the PMF of dimerization for Nbrpe-5 as the change in free energy between the large d plateau (i.e., noninteracting chains) and the minimum of the free-energy curves at $d = 1.0$ nm, we find favorable (negative) values of $\Delta F_{\text{Nbrpe}}^{\text{CG}} = -(20.7 \pm 1.6)$ kJ/mol and $\Delta F_{\text{Nbrpe}}^{\text{AA}} = -(17.0 \pm 2.8)$ kJ/mol that are in agreement to within 3.7 kJ/mol ($1.5 k_B T$ at $T = 300$ K). Employing the same definition for sarcosine-5, we observe unfavorable (positive) values of $\Delta F_{\text{sarcosine}}^{\text{CG}} = +(8.4 \pm 0.5)$ kJ/mol and $\Delta F_{\text{sarcosine}}^{\text{AA}} = +(9.4$

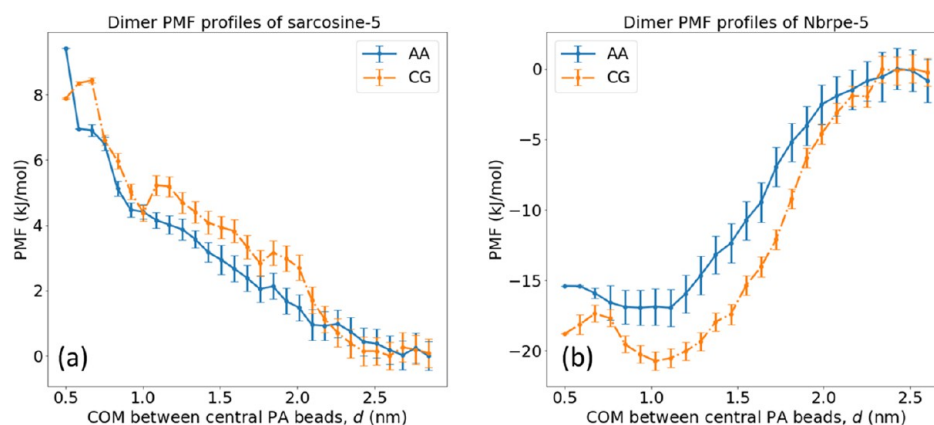


Figure 10. Dimerization PMFs for sarcosine-5 and Nbrpe-5 in water at $T = 300$ K and $P = 1$ bar. (a) Sarcosine-5 dimerization is unfavorable, whereas (b) Nbrpe-5 dimerization is spontaneous. PMF curves are computed using umbrella sampling and WHAM, and uncertainties are estimated by 200 rounds of bootstrap resampling.

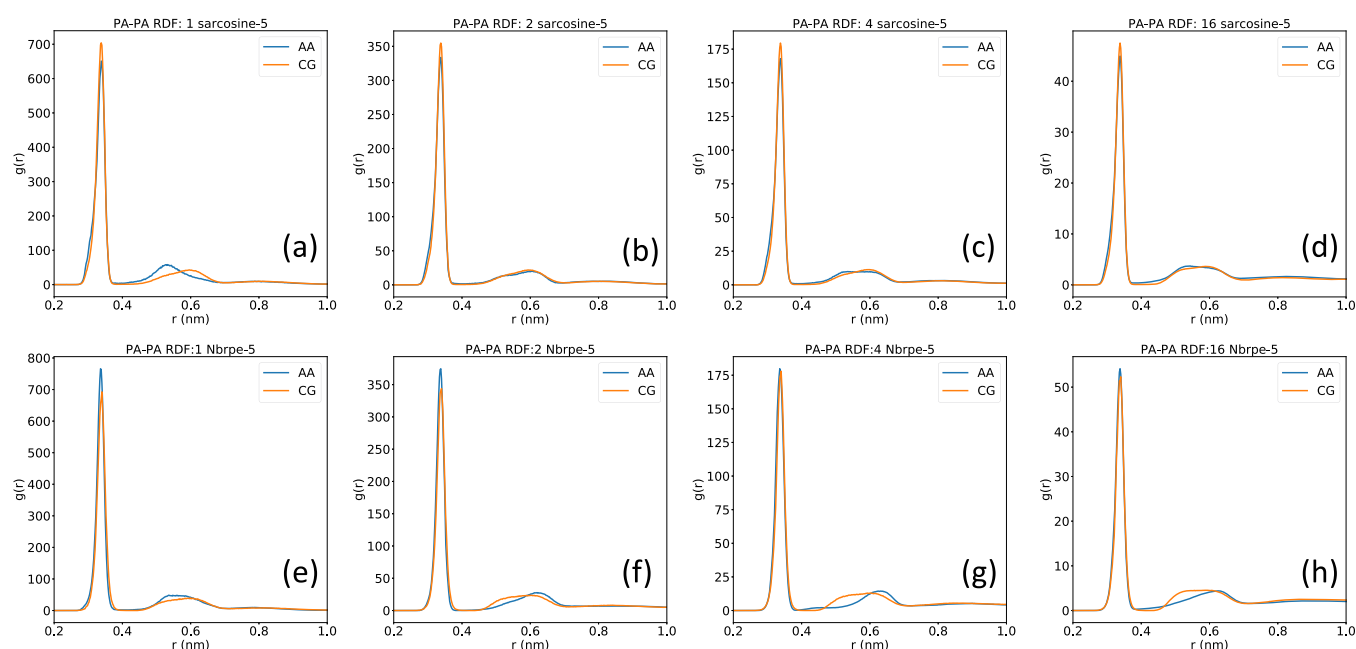


Figure 11. Backbone PA–PA radial distribution functions for polysarcosine and poly-Nbrpe in water at $T = 300$ K and $P = 1$ bar as a function of concentration. All-atom and CG simulations were performed in $4.5 \times 4.5 \times 4.5$ nm³ water boxes comprising (a, e) 1, (b, f) 2, (c, g) 4, and (d, h) 16 peptoid chains.

± 0.4) kJ/mol in excellent agreement within 1.0 kJ/mol ($0.4 k_B T$ at $T = 300$ K).

3.6.4. Backbone Radial Distribution Functions. Fourth, we probe the predictions for the time-averaged structure and thermodynamics of the peptoid backbone within multichain aggregates assessed through the radial distribution function (rdf) between the backbone PA beads. We compute the PA–PA rdf in all-atom and CG calculations comprising $N = 1, 2, 4$, and 16 chains in water boxes of $4.5 \times 4.5 \times 4.5$ nm³ at $T = 300$ K and $P = 1$ bar corresponding to a concentration range of approximately 0.02–0.3 M. The CG mapping (Figure 3) was applied to the all-atom trajectories prior to calculation of the PA–PA rdf. The rdfs show excellent agreement at all concentrations (Figure 11). The location and height of the first three peaks are visible and are clearly matched between the CG and all-atom calculations. The only significant discrepancy is a small +0.1 nm shift in the maximum of the next-nearest-neighbor peak for the single sarcosine-5 chain (Figure 11a).

3.6.5. Self-Assembly of Poly-Nbrpe. Fifth, we test the capacity of the CG model to predict the structure and dynamics of Nbrpe-5 self-assembly. We prepare 16 Nbrpe-5 chains at a concentration of 0.24 M in a $4.8 \times 4.8 \times 4.8$ nm³ water box at $T = 300$ K and $P = 1$ bar under the all-atom and CG models. The peptoids are initially randomly placed within the box, subjected to 1 ns of NPT relaxation prior to a 200 ns production run. The self-assembly pathways in the CG and all-atom systems are in good agreement. The initially dispersed monomers aggregate within ~ 50 ns into a single large cluster that subsequently ripens by ~ 100 ns into a cylindrical rod spanning the boundaries of the periodic box and which remains stable for the remainder of the simulation (Figure 12). The formation of ordered aggregates of poly-Nbrpe peptoids is consistent with the observation of elongated nanoribbon intermediates during the formation of nanosheets^{55,58} and nanotubes⁵⁶ from amphiphilic derivatives of poly-Nbrpe and poly(*N*-((4-bromophenyl)methyl)glycine) (poly-Nbrpm) peptoids.

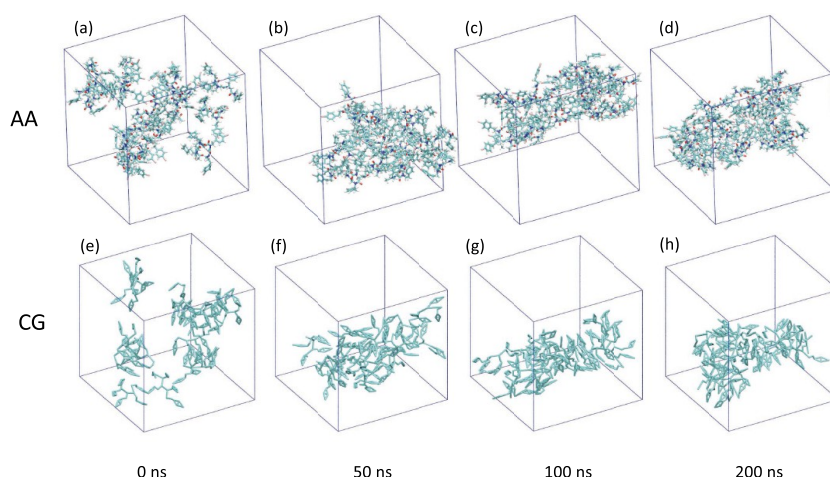


Figure 12. Self-assembly of 16 Nbrpe-5 peptoids into a cylindrical aggregate at a 0.24 M concentration in water at $T = 300$ K and $P = 1$ bar. The (a–d) all-atom and (e–h) CG simulations show similar self-assembly time courses resulting in the formation of a single cluster by ~ 50 ns that subsequently ripens by ~ 100 ns into a cylindrical rod spanning the periodic boundaries.

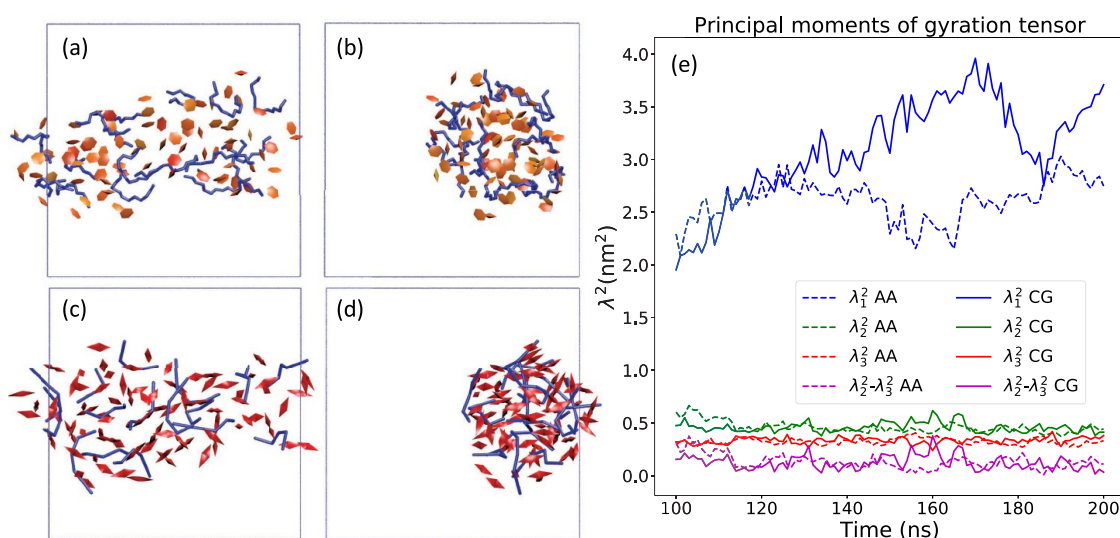


Figure 13. Quantifying the morphology of the Nbrpe-5 self-assembled cylindrical aggregate. Side and end views of the self-assembled cylindrical aggregate formed from (a, b) all-atom and (c, d) CG simulations of 16 Nbrpe-5 peptoid chains in water at $T = 300$ K and $P = 1$ bar. Peptoid backbones are represented as blue tubes and aromatic side chains as orange (all-atom) or red (CG) polygons. (e) Time evolution of the principal moments of the gyration tensor ($\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2$) and acylindricity $c = \lambda_2^2 - \lambda_3^2$ over the terminal 100 ns of the simulation.

We present side and end views of the terminal cylindrical aggregate in Figure 13a–d, where we observe good agreement in the diameter of the cylinder and relative positioning of the peptoid backbone and side chains in the all-atom and CG simulations. We quantify the morphology of the cylindrical aggregate by resolving the principal moments of the gyration tensor computed over all of the centers of mass of all atoms or beads over the terminal 100 ns of the simulation (Figure 13e).^{99,101} The mean and standard deviation of the leading principal moment λ_1^2 corresponding to the long axis of the cylindrical aggregate is in good agreement between the all-atom and CG simulations— $\lambda_1^{2(\text{AA})} = 2.61 \pm 0.20 \text{ nm}^2$ and $\lambda_1^{2(\text{CG})} = 3.10 \pm 0.47 \text{ nm}^2$ —and can be compared with the value of 1.9 nm^2 expected for an idealized cylinder with length equal to the side length of the simulation box.¹⁰² The two trailing principal moments corresponding to the short axes of the cylinder are also in excellent accord with $\lambda_2^{2(\text{AA})} = 0.46 \pm 0.06 \text{ nm}^2$ and $\lambda_2^{2(\text{CG})} = 0.46 \pm 0.05 \text{ nm}^2$, and $\lambda_3^{2(\text{AA})} = 0.33 \pm 0.03 \text{ nm}^2$ and $\lambda_3^{2(\text{CG})} = 0.33 \pm 0.03 \text{ nm}^2$. The approximate equality $\lambda_2 \approx \lambda_3$ results from the

cylindrical nature of the aggregate and is reflected in low acylindricity scores of $c^{(\text{AA})} = \lambda_2^{2(\text{AA})} - \lambda_3^{2(\text{AA})} = 0.13 \pm 0.07 \text{ nm}^2$ and $c^{(\text{CG})} = \lambda_2^{2(\text{CG})} - \lambda_3^{2(\text{CG})} = 0.12 \pm 0.07 \text{ nm}^2$. These results confirm that the all-atom and CG models predict the formation of cylindrical self-assembled aggregates of Nbrpe-5 with the same shape and radius.

4. CONCLUSIONS

In this work, we have developed a bottom-up coarse-grained peptoid force field based on and compatible with the popular suite of MARTINI force fields for peptides, lipids, carbohydrates, and solvents.²⁸ The coarse-grained model is rigorously parameterized against all-atom molecular simulations performed using a peptoid-adapted force field and employing parallel bias metadynamics to ensure good sampling of the slow cis/trans isomerization of the ω dihedrals and side-chain rotations.^{46,47} The bonded interactions were fitted using iterative Boltzmann inversion to converge dependencies between the parameterized degrees of freedom^{48–50} and the nonbonded interactions

specified by the potential of mean force matching of the peptoid chain extension curves.⁵¹ We construct models for polysarcosine (poly(*N*-methyl glycine)) and poly(*N*-((4-bromophenyl)-ethyl)glycine) (poly-Nbrpe) and demonstrate the structural and thermodynamic predictions of the coarse-grained model to be in excellent accord with all-atom calculations. The coarse-grained model is up to 25-fold more computationally efficient than the all-atom model and our parameterization protocol can be generically extended to peptoids with arbitrary side chains. The agreement in the predictions of the coarse-grained model with all-atom simulations of peptoid self-assembly is particularly encouraging, and in future work, we plan to use our model to understand and engineer the self-assembly of peptoid nanomaterials at time and length scales outside of those accessible to all-atom calculations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c04567>.

Dependence of all-atom distribution functions on chain length (PDF)

Input and force field files necessary to perform all-atom parallel bias metadynamics simulations and coarse-grained simulations of polysarcosine and poly-Nbrpe (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Andrew L. Ferguson – Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0002-8829-9726; Email: andrewferguson@uchicago.edu

Authors

Mingfei Zhao – Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States

Janani Sampath – Physical Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

Sarah Alamdari – Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States

Gillian Shen – Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States

Chun-Long Chen – Physical Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States; Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0002-5584-824X

Christopher J. Mundy – Physical Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States; Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0003-1378-5241

Jim Pfandtnr – Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; Physical Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States; orcid.org/0000-0001-6727-2957

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c04567>

Notes

The authors declare the following competing financial interest(s): A.L.F. is a consultant of Evozyne and a co-author of US Provisional Patents 62/853,919 and 62/900,420 and International Patent Application PCT/US2020/035206.

■ ACKNOWLEDGMENTS

This material is based upon the work supported by U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES), as part of the Energy Frontier Research Centers program: CSSAS—The Center for the Science of Synthesis Across Scales—under award number DE-SC0019288. Pacific Northwest National Laboratory (PNNL) is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830. This work was completed in part with resources provided by the University of Chicago Research Computing Center. The authors gratefully acknowledge computing time on the University of Chicago high-performance GPU-based cyberinfrastructure (grant no. DMR-1828629). S.A. acknowledges support from the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1762114. The authors thank Drs. Alexandre Tartakovsky and Peiyuan Gao for useful discussions.

■ REFERENCES

- (1) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; et al. Peptoids: A modular approach to drug discovery. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367–9371.
- (2) Zuckermann, R. N. Peptoid origins. *Biopolymers* **2011**, *96*, 545–555.
- (3) Rosales, A. M.; Murnen, H. K.; Zuckermann, R. N.; Segalman, R. A. Control of crystallization and melting behavior in sequence specific polypeptoids. *Macromolecules* **2010**, *43*, 5627–5636.
- (4) Rosales, A. M.; Murnen, H. K.; Kline, S. R.; Zuckermann, R. N.; Segalman, R. A. Determination of the persistence length of helical and non-helical polypeptoids in solution. *Soft Matter* **2012**, *8*, 3673–3680.
- (5) Weiser, L. J.; Santiso, E. E. Molecular modeling studies of peptoid polymers. *AIMS Mater. Sci.* **2017**, *4*, 1029–1051.
- (6) Sun, J.; Zuckermann, R. N. Peptoid polymers: A highly designable bioinspired material. *ACS Nano* **2013**, *7*, 4715–4732.
- (7) Yan, F.; Liu, L.; Walsh, T. R.; Gong, Y.; El-Khoury, P. Z.; Zhang, Y.; Zhu, Z.; De Yoreo, J. J.; Engelhard, M. H.; Zhang, X.; et al. Controlled synthesis of highly-branched plasmonic gold nanoparticles through peptoid engineering. *Nat. Commun.* **2018**, *9*, No. 2327.
- (8) Gellman, S. H. Foldamers: A manifesto. *Acc. Chem. Res.* **1998**, *31*, 173–180.
- (9) Mukherjee, S.; Zhou, G.; Michel, C.; Voeltz, V. A. Insights into peptoid helix folding cooperativity from an improved backbone potential. *J. Phys. Chem. B* **2015**, *119*, 15407–15417.
- (10) Hebert, M. L.; Shah, D. S.; Blake, P.; Turner, J. P.; Servoss, S. L. Tunable peptoid microspheres: Effects of side chain chemistry and sequence. *Org. Biomol. Chem.* **2013**, *11*, 4459–4464.
- (11) Murnen, H. K.; Rosales, A. M.; Jaworski, J. N.; Segalman, R. A.; Zuckermann, R. N. Hierarchical self-assembly of a biomimetic diblock copolypeptoid into homochiral superhelices. *J. Am. Chem. Soc.* **2010**, *132*, 16112–16119.
- (12) Sanii, B.; Kudirka, R.; Cho, A.; Venkateswaran, N.; Olivier, G. K.; Olson, A. M.; Tran, H.; Harada, R. M.; Tan, L.; Zuckermann, R. N. Shaken, not stirred: Collapsing a peptoid monolayer to produce free-floating, stable nanosheets. *J. Am. Chem. Soc.* **2011**, *133*, 20808–20815.
- (13) Nam, K. T.; Shelby, S. A.; Choi, P. H.; Marciel, A. B.; Chen, R.; Tan, L.; Chu, T. K.; Mesch, R. A.; Lee, B.-C.; Connolly, M. D.; et al. Free-floating ultrathin two-dimensional crystals from sequence-specific peptoid polymers. *Nat. Mater.* **2010**, *9*, 454–460.

- (14) Miller, S. M.; Simon, R. J.; Ng, S.; Zuckermann, R. N.; Kerr, J. M.; Moos, W. H. Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers. *Drug Dev. Res.* **1995**, *35*, 20–32.
- (15) Lee, J.; Udugamasooriya, D. G.; Lim, H. S.; Kodadek, T. Potent and selective photo-inactivation of proteins with peptoid-ruthenium conjugates. *Nat. Chem. Biol.* **2010**, *6*, 258–260.
- (16) Robertson, E. J.; Battigelli, A.; Proulx, C.; Mannige, R. V.; Haxton, T. K.; Yun, L.; Whitelam, S.; Zuckermann, R. N. Design, synthesis, assembly, and engineering of peptoid nanosheets. *Acc. Chem. Res.* **2016**, *49*, 379–389.
- (17) Luo, Y.; Song, Y.; Wang, M.; Jian, T.; Ding, S.; Mu, P.; Liao, Z.; Shi, Q.; Cai, X.; Jin, H.; et al. Bioinspired peptoid nanotubes for targeted tumor cell imaging and chemo-photodynamic therapy. *Small* **2019**, *15*, No. 1902485.
- (18) Song, Y.; Wang, M.; Li, S.; Jin, H.; Cai, X.; Du, D.; Li, H.; Chen, C.-L.; Lin, Y. Efficient cytosolic delivery using crystalline nanoflowers assembled from fluorinated peptoids. *Small* **2018**, *14*, No. 1803544.
- (19) Möhle, K.; Hofmann, H.-J. Peptides and peptoids — A quantum chemical structure comparison. *Biopolymers* **1996**, *38*, 781–790.
- (20) Möhle, K.; Hofmann, H.-J. Peptides and peptoids — A systematic structure comparison. *J. Mol. Model.* **1996**, *2*, 307–311.
- (21) Butterfoss, G. L.; Renfrew, P. D.; Kuhlman, B.; Kirshenbaum, K.; Bonneau, R. A preliminary survey of the peptoid folding landscape. *J. Am. Chem. Soc.* **2009**, *131*, 16798–16807.
- (22) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (23) Mirijanian, D. T.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *J. Comput. Chem.* **2014**, *35*, 360–370.
- (24) Weiser, L. J.; Santiso, E. E. A CGenFF-based force field for simulations of peptoids with both cis and trans peptide bonds. *J. Comput. Chem.* **2019**, *40*, 1946–1956.
- (25) Haxton, T. K.; Zuckermann, R. N.; Whitelam, S. Implicit-solvent coarse-grained simulation with a fluctuating interface reveals a molecular mechanism for peptoid monolayer buckling. *J. Chem. Theory Comput.* **2016**, *12*, 345–352.
- (26) Du, P.; Rick, S. W.; Kumar, R. Towards a coarse-grained model of the peptoid backbone: The case of N, N -dimethylacetamide. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23386–23396.
- (27) Gao, P.; Tartakovsky, A. MARTINI-Based Coarse-Grained Model for Poly(alpha-peptoid)s. 2019, arXiv:1903.01975. arXiv.org e-Print archive. <https://arxiv.org/abs/1903.01975>.
- (28) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.
- (29) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (30) Molinero, V.; Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **2009**, *113*, 4008–4016.
- (31) Stillinger, F. H.; Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **1985**, *31*, 5262–5271.
- (32) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (33) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (34) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (35) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (36) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *49*, 61–74.
- (37) Voelz, V. A.; Dill, K. A.; Chorny, I. Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Biopolymers* **2011**, *96*, 639–650.
- (38) Feigel, M. Rotation barriers of amides in the gas phase. *J. Phys. Chem. A* **1983**, *87*, 3054–3058.
- (39) Sui, Q.; Borchardt, D.; Rabenstein, D. L. Kinetics and equilibria of cis/trans isomerization of backbone amide bonds in peptoids. *J. Am. Chem. Soc.* **2007**, *129*, 12042–12048.
- (40) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press, 2003.
- (41) Weber, B.; Birke, A.; Fischer, K.; Schmidt, M.; Barz, M. Solution properties of polysarcosine: From absolute and relative molar mass determinations to complement activation. *Macromolecules* **2018**, *51*, 2653–2661.
- (42) Birke, A.; Ling, J.; Barz, M. Polysarcosine-containing copolymers: Synthesis, characterization, self-assembly, and applications. *Prog. Polym. Sci.* **2018**, *81*, 163–208.
- (43) Cui, S.; Pan, X.; Gebru, H.; Wang, X.; Liu, J.; Liu, J.; Li, Z.; Guo, K. Amphiphilic star-shaped poly (sarcosine)-block-poly (ϵ -caprolactone) diblock copolymers: one-pot synthesis, characterization, and solution properties. *J. Mater. Chem. B* **2017**, *5*, 679–690.
- (44) Luxenhofer, R.; Fetsch, C.; Grossmann, A. Polypeptoids: A perfect match for molecular definition and macromolecular engineering. *J. Polym. Sci., Part A: Polym. Chem.* **2013**, *51*, 2731–2752.
- (45) Klinker, K.; Barz, M. Polypept(o)ides: Hybrid systems based on polypeptides and polypeptoids. *Macromol. Rapid Commun.* **2015**, *36*, 1943–1957.
- (46) Pfaendtner, J.; Bonomi, M. Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 5062–5067.
- (47) Prakash, A.; Fu, C. D.; Bonomi, M.; Pfaendtner, J. Biasing smarter, not harder, by partitioning collective variables into families in parallel bias metadynamics. *J. Chem. Theory Comput.* **2018**, *14*, 4985–4990.
- (48) Müller-Plathe, F. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem* **2002**, *3*, 754–769.
- (49) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.
- (50) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, No. 090901.
- (51) Wang, J.; Ferguson, A. L. Mesoscale simulation of asphaltene aggregation. *J. Phys. Chem. B* **2016**, *120*, 8016–8035.
- (52) Sigmund, F.; Wessely, F. Untersuchungen über α -Amino-N-Carbonsäureanhydride. II. *Hoppe-Seyler's Z. Physiol. Chem.* **1926**, *157*, 91–105.
- (53) Lau, K. H. A.; Ren, C.; Sileika, T. S.; Park, S. H.; Szleifer, I.; Messersmith, P. B. Surface-grafted polysarcosine as a peptoid antifouling polymer brush. *Langmuir* **2012**, *28*, 16099–16107.
- (54) Gangloff, N.; Ulbricht, J.; Lorson, T.; Schlaad, H.; Luxenhofer, R. Peptoids and polypeptoids at the frontier of supra- and macromolecular engineering. *Chem. Rev.* **2016**, *116*, 1753–1802.
- (55) Jin, H.; Jiao, F.; Daily, M. D.; Chen, Y.; Yan, F.; Ding, Y. H.; Zhang, X.; Robertson, E. J.; Baer, M. D.; Chen, C. L. Highly stable and self-repairing membrane-mimetic 2D nanomaterials assembled from lipid-like peptoids. *Nat. Commun.* **2016**, *7*, No. 12252.
- (56) Jin, H.; Ding, Y. H.; Wang, M.; Song, Y.; Liao, Z.; Newcomb, C. J.; Wu, X.; Tang, X. Q.; Li, Z.; Lin, Y.; et al. Designable and dynamic single-walled stiff nanotubes assembled from sequence-defined peptoids. *Nat. Commun.* **2018**, *9*, No. 270.

- (57) Jiao, F.; Chen, Y.; Jin, H.; He, P.; Chen, C. L.; De Yoreo, J. J. Self-repair and patterning of 2D membrane-like peptoid materials. *Adv. Funct. Mater.* **2016**, *26*, 8960–8967.
- (58) Xuan, S.; Jiang, X.; Spencer, R. K.; Li, N. K.; Prendergast, D.; Balsara, N. P.; Zuckermann, R. N. Atomic-level engineering and imaging of polypeptoid crystal lattices. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 22491–22499.
- (59) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *Solving Software Challenges for Exascale. EASC 2014*; Markidis, S.; Laure, E., Eds.; Lecture Notes in Computer Science; Springer, 2015; pp 3–27.
- (60) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (61) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminf.* **2012**, *4*, No. 17.
- (62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (63) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
- (64) Berendsen, H. J.; Postma, J. P.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (65) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (66) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (67) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press, 1988.
- (68) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (69) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (70) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (71) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (72) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (73) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.
- (74) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (75) Hoyas, S.; Lemaire, V.; Duez, Q.; Saintmont, F.; Halin, E.; De Winter, J.; Gerbaux, P.; Cornil, J. PEPDROID: Development of a generic DREIDING-based force field for the assessment of peptoid secondary structures. *Adv. Theory Simul.* **2018**, *1*, No. 1800089.
- (76) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (77) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (78) ChemDraw, version 18.2.0.37; PerkinElmer Informatics, Inc., 1985–2019. <https://www.perkinelmer.com.cn/category/chemdraw>.
- (79) Alessandri, R.; Uusitalo, J. J.; De Vries, A. H.; Havenith, R. W.; Marrink, S. J. Bulk heterojunction morphologies with atomistic resolution from coarse-grain solvent evaporation simulations. *J. Am. Chem. Soc.* **2017**, *139*, 3697–3705.
- (80) Lindahl, E.; Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B. Implementation of the charmm force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.
- (81) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- (82) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.
- (83) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 2017.
- (84) Villa, A.; van der Vegt, N. F.; Peter, C. Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2068–2076.
- (85) Villa, A.; Peter, C.; van der Vegt, N. F. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2077–2086.
- (86) Mansbach, R. A.; Ferguson, A. L. Coarse-grained molecular simulation of the hierarchical self-assembly of π -conjugated optoelectronic peptides. *J. Phys. Chem. B* **2017**, *121*, 1684–1706.
- (87) Thurston, B. A.; Tovar, J. D.; Ferguson, A. L. Thermodynamics, morphology, and kinetics of early-stage self-assembly of π -conjugated oligopeptides. *Mol. Simul.* **2016**, *42*, 955–975.
- (88) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (89) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. g-wham: A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.
- (90) Dalgicdir, C.; Sensoy, O.; Peter, C.; Sayar, M. A transferable coarse-grained model for diphenylalanine: How to represent an environment driven conformational transition. *J. Chem. Phys.* **2013**, *139*, No. 234115.
- (91) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.
- (92) Mashayak, S. Y.; Jochum, M. N.; Koschke, K.; Aluru, N. R.; Rühle, V.; Junghans, C. Relative entropy and optimization-driven coarse-graining methods in VOTCA. *PLoS One* **2015**, *10*, No. e0131754.
- (93) Ryckaert, J.-P.; Bellemans, A. Molecular dynamics of liquid alkanes. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95–106.
- (94) de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (95) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (96) Groot, R. D. Electrostatic interactions in dissipative particle dynamics—simulation of polyelectrolytes and anionic surfactants. *J. Chem. Phys.* **2003**, *118*, 11265–11277.
- (97) Groot, R. D.; Rabone, K. Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. *Biophys. J.* **2001**, *81*, 725–736.
- (98) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier, 2001.
- (99) Theodorou, D. N.; Suter, U. W. Shape of unperturbed linear polymers: Polypropylene. *Macromolecules* **1985**, *18*, 1206–1214.
- (100) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (101) Mattice, W. L.; Suter, U. W. *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems*; Wiley-Interscience, 1994.

(102) Mansfield, M. L.; Douglas, J. F. Shape characteristics of equilibrium and non-equilibrium fractal clusters. *J. Chem. Phys.* **2013**, *139*, No. 044901.