The Journal of Chemical Physics

A hybrid approach for coarse-graining helical peptoids: Solvation, secondary structure, and assembly

Cite as: J. Chem. Phys. **158**, 114105 (2023); https://doi.org/10.1063/5.0138510
Submitted: 12 December 2022 • Accepted: 21 February 2023 • Accepted Manuscript Online: 22 February 2023 • Published Online: 15 March 2023

Akash Banerjee and Meenakshi Dutt

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

A hybrid approach for coarse-graining helical peptoids: Solvation, secondary structure, and assembly

Cite as: J. Chem. Phys. 158, 114105 (2023); doi: 10.1063/5.0138510
Submitted: 12 December 2022 • Accepted: 21 February 2023 •
Published Online: 15 March 2023

Akash Banerjee and Meenakshi Dutt[a)]

AFFILIATIONS

Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA

[a)]Author to whom correspondence should be addressed: meenakshi.dutt@rutgers.edu

ABSTRACT

Protein mimics such as peptoids form self-assembled nanostructures whose shape and function are governed by the side chain chemistry and secondary structure. Experiments have shown that a peptoid sequence with a helical secondary structure assembles into microspheres that are stable under various conditions. The conformation and organization of the peptoids within the assemblies remains unknown and is elucidated in this study via a hybrid, bottom-up coarse-graining approach. The resultant coarse-grained (CG) model preserves the chemical and structural details that are critical for capturing the secondary structure of the peptoid. The CG model accurately captures the overall conformation and solvation of the peptoids in an aqueous solution. Furthermore, the model resolves the assembly of multiple peptoids into a hemispherical aggregate that is in qualitative agreement with the corresponding results from experiments. The mildly hydrophilic peptoid residues are placed along the curved interface of the aggregate. The composition of the residues on the exterior of the aggregate is determined by two conformations adopted by the peptoid chains. Hence, the CG model simultaneously captures sequence-specific features and the assembly of a large number of peptoids. This multiscale, multiresolution coarse-graining approach could help in predicting the organization and packing of other tunable oligomeric sequences of relevance to biomedicine and electronics.

I. INTRODUCTION

Nature's critical processes rely on numerous proteins working in tandem. The properties and functions of these proteins are dictated by the sequence, structure, and organization of their subunits. However, these proteins are unable to sustain their functionality upon exposure to proteases and changes in temperature, pH, and non-biological conditions. These shortcomings can be overcome by synthetic heteropolymers that mimic the design principles and properties and sustain the functionalities of proteins under a variety of external conditions. Synthetic heteropolymers constituting protein-mimics encompass higher order structures corresponding to those of the protein subunits. The subunits include secondary structures such as helices, beta-sheets, their complexes, or tertiary structures.[1,2] The sequence, structure, packing, and organization of the subunits are key to the function of the synthetic heteropolymer. Synthetic polymers with side chain diversity and secondary structure have been synthesized toward the goal of adopting stable tertiary and quaternary structures and displaying protein-like functions.[3,4] Yet, most synthetic polymers have a limited chemical diversity or monomer coupling reaction efficiency. However, a class of synthetic polymers, namely, poly-N-substituted glycine heteropolymers or peptoids,[5,6] are not constrained by these limitations and adopt higher order structures reminiscent of protein subunits.

Unlike proteins and peptides where the side chain is appended to the backbone α-carbon, the side chains in peptoid sequences are appended to the backbone nitrogen.[7,8] This modification results in a loss of chirality and hydrogen bonding along the peptoid backbone, i.e., the key features that determine the secondary structure of proteins and peptides. In peptoids, the secondary structure is determined by the side chains. As a consequence, peptoids are able to sustain their secondary structure upon exposure to proteases and changes in temperature, pH, and non-biological conditions. Furthermore, the side chain diversity in peptoids is expansive. These characteristics enable the design of highly tunable sequences that aggregate to form a variety of ordered assemblies, including

superhelices,[9] nanosheets,[10,11] nanotubes,[12] and vesicles.[13] The secondary structure of individual peptoid chains influences the structure and morphology of the final equilibrated aggregate.[11] However, the preservation of the secondary structure along with the molecular organization upon the formation of a thermodynamically stable aggregate remains an open question. Of specific interest are peptoid sequences with a helical secondary structure. This study focuses on a sequence that assembles into microspheres.[14,15]

A prior experimental study demonstrated a helical peptoid sequence with a chiral, aromatic side chain to self-assemble into microspheres.[14,15] The side chain residues on the third face of the helix were tuned to study their impact on the viability and characteristics of the microspheres. Secondary structure and partial water solubility were identified as key features that determined the robustness of the microspheres. However, due to the constraints in experimental techniques, the organization of the peptoids within microspheres remains unknown. On the other hand, computational approaches can resolve detailed characteristics of the molecules within microspheres. To date, existing computational models have characterized the properties of a single peptoid helix.[16,17] These models are not suitable for capturing the self-assembly of peptoids into larger aggregates. This challenge can be addressed by a computational technique that simultaneously captures the assembly over a range of spatiotemporal scales and resolves the secondary structure and conformation of the molecules within the aggregates.

In this study, the self-assembly of a peptoid sequence with a helical secondary structure is investigated using a multiscale, multiresolution computational approach. The computational approach uses a multiresolution coarse-grained (CG) model in conjunction with the Molecular Dynamics (MD) simulation technique to resolve the dynamics of the self-assembly process. The CG model of the peptoid sequence is developed to resolve the solvation, structure of individual molecules, and their assembly. A bottom-up model development approach is employed, wherein the CG model is parameterized using all atom (AA) MD trajectories. This CG model also provides insights into the molecular organization of the peptoid sequences within an aggregate. Due to the bottom-up nature of this CG model, the AA characteristics are preserved. This allows one to reintroduce the atomistic representation in the CG configurations via backmapping. Hence, the term "multiresolution" is used for this class of bottom-up CG approaches.[18–20]

All atom models represent the detailed chemical structure of oligomeric sequences of peptoids. Due to the modifications along the backbone, the frequently used peptide force fields cannot accurately capture features such as cis–trans isomerization in peptoids.[21] Hence, peptide force fields were modified to build dedicated AA peptoid force fields, such as MFTOID.[22] Based on the experimental and quantum mechanical data, seven parameters in the CHARMM22 force field were modified to obtain the MFTOID force field. Another study[23] employed the Chemistry at Harvard Macromolecular Mechanics (CHARMM) generalized force field (CGenFF)[24] to refine the MFTOID force field. This approach yielded better results for three peptoid residues and provided a protocol for transferability to any arbitrary peptoid residue. Other peptide-based[25] and generic[16,26] force fields have been modified to create AA peptoid force fields. Overall, the main concern with AA peptoid force fields is transferability across a wide range of side chain residues. To that end, MFTOID has been tested on various side chain residues. It provides good results for simulations with single[27] and multiple peptoid[28] chains. A previous study demonstrated that MFTOID was able to resolve the assembly of multiple peptoids into a nanosheet with dimensions that are in agreement with experiments.[11] In addition, the model was able to resolve a secondary structure that is unique to the specific peptoid sequence. Furthermore, MFTOID was employed to resolve the assembly of a large number of peptoids into helical rods and crystalline sheets.[28] These AA simulations began with a preassembled structure wherein the peptoids were placed close to each other.[29] It is not computationally feasible to resolve the assembly of a large number of randomly distributed peptoids into an aggregate that is at thermodynamic equilibrium. Hence, although AA models preserve the chemical details of the peptoid sequences, these models are unable to capture their self-assembly due to the enormous degrees of freedom and the extended spatiotemporal scales.

To resolve the assembly of peptoids over large spatiotemporal scales, CG models are employed. These models represent the peptoid sequences with significantly fewer degrees of freedom and can be categorized as top-down or bottom-up models. The top-down CG models[30] are parameterized using thermodynamic properties, such as the experimental free-energy of partitioning, and are typically transferable. For example, the MARTINI force field[31,32] efficiently resolves the assembly of lipids and peptides while simplifying the representation of the chemical details of individual molecules, such as the secondary structure. In addition, the top-down CG models can be used to access large spatiotemporal scales. This is attributed to the simple and smooth nature of the underlying CG potentials. These potentials are fitted to an analytical function, and they typically have a single energy well. In addition, the top-down approaches do not employ small CG beads. On average, the MARTINI model employs a 4:1 mapping scheme, where four heavy atoms (such as carbon and oxygen) are mapped to a single CG bead. These features enable the use of a large integration time step. For example, an integration time step between 20 and 30 fs can be used with the MARTINI model.[33] Currently, there are no top-down CG models available for peptoids. In contrast, the bottom-up CG models[30,34] are parameterized using trajectories from AA MD simulations. Typically, the coarse-graining scheme of bottom-up approaches for oligomeric sequences, such as peptoids and peptides, include small CG beads.[35,36] Such beads are required to accurately resolve the chemical structure of the sequence. The bottom-up approaches that include small CG beads employ a small integration time step (2 fs). The use of larger integration time steps (such as 4 fs) leads to numerical instability. As suggested by a previous study,[35] a smaller integration time step could be required to maintain numerical stability when there are small CG beads in the model. The previous study employed a 2 fs time step for a peptoid sequence that had bulky side chains like the ones considered in this study. The coarse-graining of bulky side chains requires small CG beads that encompass atomistic fragments such as CH_2 groups (see Fig. 1). Another bottom-up study on peptides[36] also employed a time step of 2 fs. Due to a smaller integration time step, these bottom-up CG models were unable to access large spatiotemporal scales.

Bottom-up CG models that resolve the assembly of peptoids can be categorized based on the starting configuration of the simulation. Some studies model peptoid-based assemblies that start from a preassembled initial configuration. In these studies, the peptoids are placed in close proximity and they aggregate to form a large

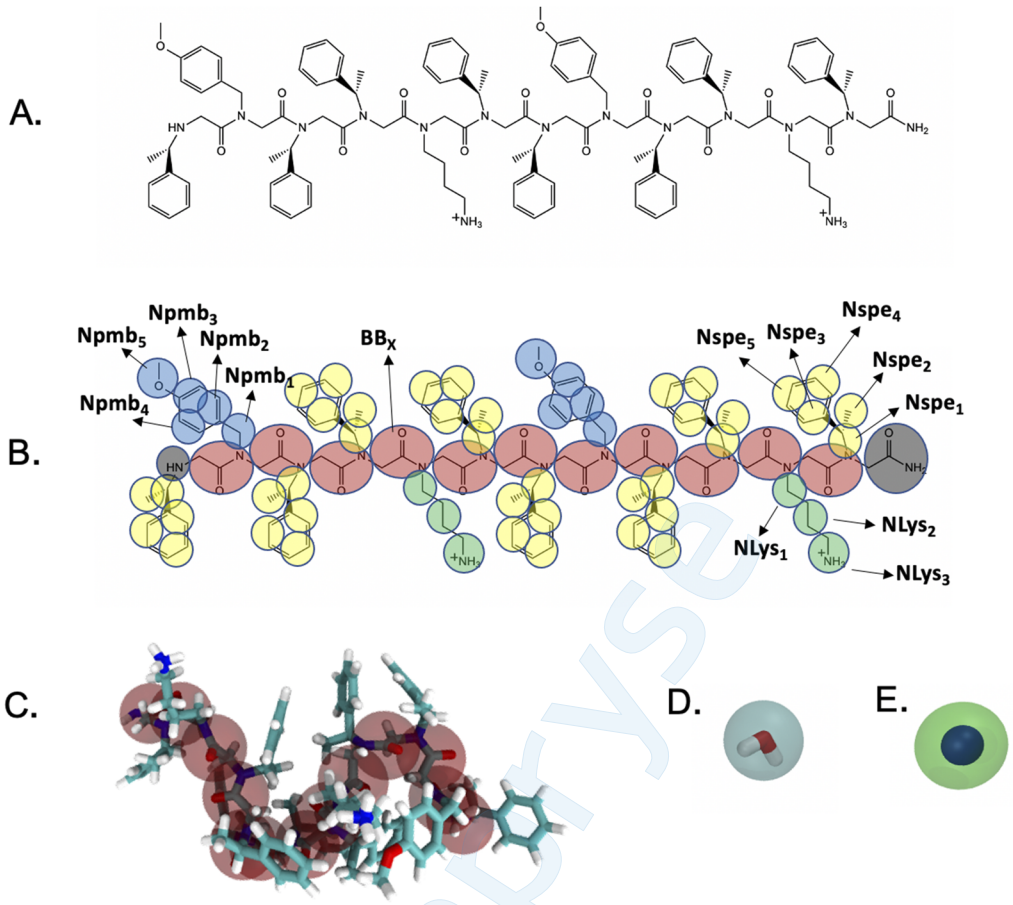

FIG. 1. (a) The [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence. (b) CG scheme superimposed on the peptoid sequence. The red beads represent the peptoid backbone. The gray beads represent the C and N terminal beads. The yellow, blue, and green beads represent the Nspe, Npmb, and NLys peptoid residues, respectively. Zoomed-in views of these residues are shown in Fig. SI.2. (c) Relaxed helix-like AA configuration of the peptoid sequence. The backbone beads are superimposed on the AA coordinates for ease of visualization of the helix-like structure. (d) and (e) CG schemes for water molecules and ions, respectively. For ions, the CG bead is shown larger for ease of visualization.

assembly over the course of the simulation. An example of this approach is the MF-CG-TOID force field, which was employed to study the formation of nanosheets using anisotropic CG potentials.[37] The CG model was fitted to trajectories from AA simulations (using the MFTOID force field) and further refined using experimental data. The model resolved the assembly of 48 peptoids into a monolayer on an air–water interface. Due to the bottom-up nature of the model, the average spacing between the peptoid chains was in quantitative agreement with the x-ray scattering data. However, it is noted that these simulations started from a configuration wherein the peptoids were preassembled in a brick-like arrangement. The issue with preassembled models is that the conformation of individual peptoids could be biased. Since the peptoids are placed in close proximity, the conformational sampling of individual peptoid chains could be limited. This could result in biased conformations of individual peptoid chains. Other investigations model the self-assembly of a few peptoids starting from a randomly distributed initial configuration. Here, the peptoids are initially placed outside the interaction range from each other. Hence, the initial conformation of the peptoids is unbiased. Currently, to the best of our knowledge, there is only one such work that models the assembly of 16 CG peptoids.[35] In this model, the CG time scale is nearly the same as the AA time scale. Hence, with this CG model, it could be computationally expensive to resolve the assembly of a larger number of peptoids (namely, of the order of 100 or more molecules). Therefore, there is a need for a CG model that captures the self-assembly of a large number of peptoids. Initially, the peptoids need to be randomly distributed and placed outside the interaction range from each other. This will ensure that the initial configuration of the peptoids is unbiased. Hence, this study reports a model that captures the self-assembly of 128 peptoids starting from a randomly distributed initial configuration.

This study will develop a multiresolution CG model that simultaneously captures the solvation, structure, and assembly of peptoid sequences with a helical secondary structure. This can be achieved by a combination of bottom-up CG techniques as demonstrated in an earlier study.[20] Typically, AA trajectories provide the structure,[38,39]

forces,[40] and energies[19,35,36,41] of the molecules, which can be used to develop the CG potentials using various coarse-graining methods. In this study, a combination of structure and force-based methods are used. The bonded interactions, i.e., bonds, angles, and dihedrals, are derived using a structure-based method called Iterative Boltzmann Inversion (IBI).[38,39] This technique preserves the structure of the peptoid in the CG model. In addition, IBI is used to resolve the solvation of the peptoid, i.e., the peptoid–water interactions. To resolve the assembly of the peptoids, the intermolecular interactions between the peptoids need to be sampled to develop the CG potentials. An earlier study[20] demonstrated IBI to generate multi-well peptide–peptide CG potentials. Since these potentials are highly coupled to the structure of the reference system, the transferability of these potentials may be limited. Hence, a force-based method called Force Matching (FM)[40,42] is used to resolve the peptoid–peptoid interactions. This technique generates single-well potentials that are expected to be transferable over a limited range of concentrations of the peptoids. Since two coarse-graining techniques are employed to build the CG model (IBI and FM), the current method is termed as a "hybrid," bottom-up coarse-graining approach. The term "hybrid" is used to be consistent with a prior study that employed IBI and FM to derive a CG model.[42] The CG force field is tested for transferability across a peptoid concentration range. Hence, in this work, the term "transferability" is only used in the context of concentration.

In this study, a variation of an existing bottom-up coarse-graining approach[42,43] is employed. The existing approach uses the Boltzmann inversion and the FM technique. It has been used for developing CG force fields to study protein folding, conformations, and association.[43,44] The study on protein association[43] examined the association of five alanine-15 sequences in a simulation box with a focus on the CG force development and not on the aggregation process, the characteristics of the assembly, the molecular packing, or organization within the assemblies. In this study, the IBI and FM techniques are used to develop a bottom-up CG force field that is able to capture the aggregation of helical peptoids, study the characteristics of the assembly, and probe the molecular conformation, packing, and organization within the assemblies. Furthermore, the CG model is used to study systems sizes that are 1–2 orders of magnitude larger than what has been investigated to date.[43]

The hybrid, bottom-up coarse-graining approach is employed to develop a CG model for the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence in an aqueous solution. The helical conformation of the peptoid sequence is preserved in the CG model. In addition, the solvation structure of the peptoid is in agreement with corresponding results from AA simulations. The CG model resolves the assembly of multiple peptoids into a hemispherical aggregate that is in qualitative agreement with corresponding results from experiments.[14,15] This hybrid, bottom-up coarse-graining approach can be extended to other oligomeric sequences whose molecular structure, organization, and assembly need to be precisely resolved.

This work develops a fundamental understanding of how helical peptoids pack within assemblies. A CG model that preserves the helical secondary structure of individual peptoid chains and captures the assembly of the peptoids is reported. The self-assembly of 64 CG peptoids yields a hemispherical aggregate. Furthermore, the self-assembly of 128 CG peptoids yields an assembly that consists of two connected hemispherical aggregates. Hence, it is surmised that larger peptoid-based assemblies may consist of a network of hemispherical aggregates. Since the side chain residues mainly govern the behavior of peptoid-based materials, their organization within the hemispherical aggregate is characterized. The peptoid sequence consists of three types of side chains: (a) charged, hydrophilic, (b) mildly hydrophilic, and (c) hydrophobic residues. The mildly hydrophilic residues are placed along the curved interface of the hemispherical aggregate. The helical peptoids adopt two conformations within the aggregate. These conformations dictate the placement of the charged, hydrophilic residues on the exterior of the aggregate. Finally, the hydrophobic residues mainly drive the self-assembly of the peptoid chains into an aggregate. This information can be used to guide the future design and synthesis of tunable materials for use in a diverse range of disciplines, including biomedicine,[45–47] energy,[48,49] and electronics.[50,51]

II. METHODS

All simulation files are provided in Ref. 52. In the supplementary material, Fig. SI.1 lists the computational tools provided in Ref. 52. These files can be employed to reproduce all parts of this work. The AA and CG simulation details are provided in the supplementary material.

A. Mapping scheme

A coarse-grained model for the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence (Fig. 1) is developed. This sequence consists of three peptoid residues, namely, Nspe [*N*-(*S*)-1-phenylethylglycine], Npmb [*N*-(*p*-methoxybenzyl)glycine], and NLys (*N*-4-aminobutylglycine). The backbone and side chain residues of the peptoid are coarse-grained into separate beads. The peptoid sequence consists of 12 sequentially connected residues. Eleven coarse-grained beads are employed to represent the peptoid backbone. These beads are referred to as BB_X (BB: backbone; X: index of the backbone residue). Each backbone bead is connected to a side chain residue. The Nspe and Npmb side chains are coarse-grained into five beads, i.e., $Nspe_Y$ and $Npmb_Y$, respectively (Y: index of the side chain bead), whereas NLys is coarse-grained into three beads, i.e., $NLys_Y$ (Y: index of the side chain bead). Both the C and N termini are represented by a single coarse-grained bead. In addition to peptoid molecules, the system encompasses water molecules and monovalent chloride ions, which are included in the coarse-grained representation using a single-site mapping scheme.[20] It is noted that the corresponding experiments[15] are conducted in a 4:1 mixture of ethanol and water, respectively. To simplify the model development process, a single component solvent system, i.e., pure water, is chosen. In the future, multi-component solvent systems (ethanol–water mixtures) will be considered to better reproduce experimental conditions. Another CG model for peptoids has employed a multi-component representation of the solvent.[53]

B. All atom reference for coarse-grained model development

The MFTOID force field[22] is used to obtain the AA MD trajectories for the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence. A previous study compared the results for end-to-end distributions of a polypeptoid obtained from AA simulations and experiments.[27] The results for the AA representation were obtained

using four AA peptoid force fields. The MFTOID[22] and GAFF-ϕ[16] force fields yielded end-to-end distributions that were in good agreement with experiments. Since the GAFF-ϕ force field was specifically designed for Nspe residues, it imposed a bias on the ϕ dihedral. This force field could work for other chiral side chains, but could potentially yield inaccurate representations of achiral side chains. Since the current study involves both chiral and achiral peptoid residues, the MFTOID force field is chosen. The MFTOID force field is not transferable to all types of peptoid side chain residues.[27,54,55] However, the side chains that are considered in this study have been previously studied using MFTOID.[35,56] In addition, the resulting dihedral distributions for the three side chains (generated using the MFTOID force field) are validated with quantum mechanical calculations (discussed later). The MFTOID force field is a modified version of the CHARMM22 force field[57,58] and enables cis–trans isomerization by enhancing the flexibility of the ω dihedral (see Fig. SI.3). This backbone dihedral defines the isomeric states. The cis and trans states correspond to ω = 0 and ω = ±180, respectively. Figure SI.4 shows that a standard AA MD simulation cannot explore both cis and trans isomeric states within time scales that are computationally feasible. Previous studies report an energy barrier of 25 k_BT (T = 300 K) between cis and trans isomers.[35,55,59,60] Hence, parallel-biased metadynamics[35,61,62] is used in conjunction with the AA MD simulations. This enhanced sampling technique overcomes the energy barrier and samples both states. Since the method can simultaneously apply multiple 1-dimensional potentials, it performs well in simulations that encompass a large number of ω dihedrals, such as the ones used in this study. As discussed in an earlier study,[35] this method is an improvement over well-tempered metadynamics[63] that does not scale well for polypeptoids having more than 2–3 residues. Since the peptoid sequence in this study has 12 residues, well-tempered metadynamics is not considered as a suitable enhanced sampling technique. Hence, parallel-biased metadynamics, which scales well for long polypeptoids, is chosen for computational efficiency. The input parameters for parallel-biased metadynamics are the same as those adopted by an earlier study.[35] Figure SI.4 shows that the AA MD simulation used in conjunction with parallel-biased metadynamics (see the red dashed lines) samples both cis and trans isomers. The final trajectory is unbiased and is used to evaluate the energy barrier between cis and trans states. Figure SI.5 shows that the cis–trans energy barrier is in agreement with the corresponding results from earlier studies on peptoids.[22,35,55,59,60,64] Additional testing of the AA force field is performed on sarcosine, wherein the cis–trans energy barrier is in agreement with previous measurements.[22,35]

Parallel-biased metadynamics is also used on the Ψ, Φ, and ρ dihedrals (see Fig. SI.3) to sample all possible backbone and side chain conformations. Figure SI.6 shows the energy landscape of the peptoid backbone as a function of Ψ, ω and Φ, ω combinations. These results are in good agreement with prior experimental and computational studies.[54]

Figure SI.7 shows the distributions for the χ_1 dihedral for the three peptoid residues. The peak positions of these distributions are validated against corresponding results from other studies.[22,23] To enable comparison with the literature, these measurements are performed on single peptoid residues *in lieu* of the complete [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ sequence. A quantum mechanical approach for parameterizing the AA side chain parameters can also be employed for better results.[23] This approach will be pursued in the future.

C. Coarse-grained potential development

The Boltzmann Inversion (BI) and IBI[38,39] methods are employed to derive the bonded, peptoid–water, and water–water potentials. These coarse graining methods use a reference, such as the radial distribution function (RDF) of a pairwise interaction. Since the RDF represents the underlying structure of the interaction, BI and IBI are called structure-based coarse graining methods.[30,36,41] Based on the mapping scheme, the CG coordinates are mapped onto the coordinates of the AA trajectories. The reference distributions are generated from these CG-mapped coordinates. Extensive sampling of the AA simulation trajectories ensures that the reference distributions are smooth. This is a critical requirement for the development of the CG potentials. The BI and IBI methods convert the reference distributions into CG potentials. First, the BI method [Eqs. (1)–(4)] is employed to generate the CG potentials. Next, if required, the IBI method is employed to refine the CG potentials. This results in a better agreement between the distributions generated from the AA and CG trajectories,

$$U^{CG}(d,T) = -k_BTln\left(\frac{P^{CG}(d,T)}{d^2}\right), \tag{1}$$

$$U^{CG}(\theta,T) = -k_BTln\left(\frac{P^{CG}(\theta,T)}{\sin(\theta)}\right), \tag{2}$$

$$U^{CG}(\phi,T) = -k_BTln\left(P^{CG}(\phi,T)\right), \tag{3}$$

$$U^{CG}(r,T) = -k_BTln(g(r,T)). \tag{4}$$

The BI method generates bonded potentials using Eqs. (1)–(3). P^{CG} refers to the normalized distribution of a bonded interaction as a function of degree of freedom (d: bond distance, θ: angle, and ϕ: dihedral angle). Similarly, BI is employed to develop nonbonded potentials using Eq. (4). g(r) is the RDF of the nonbonded interaction between two CG beads (r is the distance between the centers of mass of the two CG beads). P^{CG} and g(r) are extracted from the AA MD trajectories sampled over extended intervals of time. The resultant CG potentials are only valid at the temperature (T) at which the AA system is simulated. Hence, these potentials are state dependent. k_B is the Boltzmann constant.

The BI method performs well on molecules that do not have any correlated degrees of freedom along the backbone that result in secondary structures (such as polymers).[18,20] The method assumes that (a) the bonded and nonbonded potentials can be treated separately [see Eq. (5), where R^N represents the position of all CG beads] and (b) all bonded degrees of freedom are independent of each other [Eq. (6)]. As demonstrated in an earlier study,[20] the BI method does not work well for peptides that have various correlated degrees of freedom. Similarly, these assumptions are not valid for peptoids. The IBI method [Eq. (7)] is employed to correct the inconsistencies caused by these assumptions. This corrective scheme is applied for bonded, peptoid–water, and water–water interactions. Furthermore, special bonded potentials are added to enhance the agreement

between the results from the AA and CG trajectories (discussed in RESULTS AND DISCUSSIONS),

$$U_{Total}(R^N) = U_{Bonded}(R^N) + U_{Nonbonded}(R^N), \quad (5)$$

$$P^{CG}(d,\theta,\phi,T) = P^{CG}(d,T) \times P^{CG}(\theta,T) \times P^{CG}(\phi,T). \quad (6)$$

The IBI method [Eq. (7)] improves the potentials generated by the BI method. ρ represents the bonded (d, θ, and ϕ) and nonbonded (r) degrees of freedom. This scheme iteratively corrects potentials using the difference between the potential of the mean force of the CG distribution at the i-th step (P_i) and the reference distribution (P^{ref}). The process is repeated until there is convergence between P_i and P^{ref} distributions,

$$V_{i+1}(\rho) = V_i(\rho) + k_B T \ln\left[\frac{P_i(\rho,T)}{P^{ref}(\rho,T)}\right]. \quad (7)$$

The Force Matching (FM) method [Eq. (8)] is used to resolve the peptoid–peptoid nonbonded interactions. This method projects the forces in the AA representation onto the CG representation of the peptoids,

$$\chi^2 = \sum_m^M \sum_n^N \left| F_{mn}^{ref} - F_{mn}^{CG} \right|^2. \quad (8)$$

FM employs a least square approach to minimize the difference between the forces in the AA (F^{ref}) and CG (F^{CG}) systems. M refers to the MD coordinate frames, and N refers to the total number of CG beads. F_{mn}^{ref} denotes the forces in the AA simulation that are acting on the n-th CG bead in the m-th MD coordinate frame. F_{mn}^{CG} refers to the corresponding forces in the CG simulation. FM converts Eq. (8) into a set of linear equations. These equations can be solved once they are overdetermined. This means that the number of parameters needs to be less than the product of the total number of CG beads (N) and the MD coordinate frames (M). Hence, a large number of MD coordinate frames are sampled from the AA trajectory.

In this study, the FM with exclusions method[42] is applied to exclude the effects of bonded and intramolecular nonbonded interactions. Forces are extracted from the final 200 of a 1000 ns AA MD simulation. A modified topology is generated to exclude the forces attributed to the bonded and intramolecular nonbonded interactions. For ease of processing, the MD trajectories are decomposed into smaller blocks, each consisting of 500 frames. An average of 30 splines is employed to generate smooth peptoid–peptoid potentials.

The Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA)[65] package is employed to implement the BI, IBI, and FM methods. All CG potentials are generated in a tabulated format. These potentials account for all underlying interactions in the AA simulations, such as van der Waals and electrostatics. The development of dedicated potentials for modeling electrostatics will be pursued in the future.

It is noted that the CG force field is only applicable to the peptoid under consideration in this study. This is because the CG model is parameterized using AA trajectories of a specific peptoid sequence and, hence, is not transferable to other peptoid sequences. Therefore, for new peptoid sequences, the entire coarse-graining procedure needs to be repeated. Following an earlier work on peptides,[43] an approach to build a transferable CG force field for peptoids will be pursued in the future. This will involve sampling AA MD simulation trajectories of multiple peptoid sequences.

III. RESULTS AND DISCUSSIONS

A systematic coarse graining approach is employed to resolve the self-assembly of peptoids into ordered aggregates while preserving their molecular structure. AA trajectories of peptoids in aqueous solution serve as a reference for the development of the CG model. The AA simulations reproduce the helical conformation of the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence

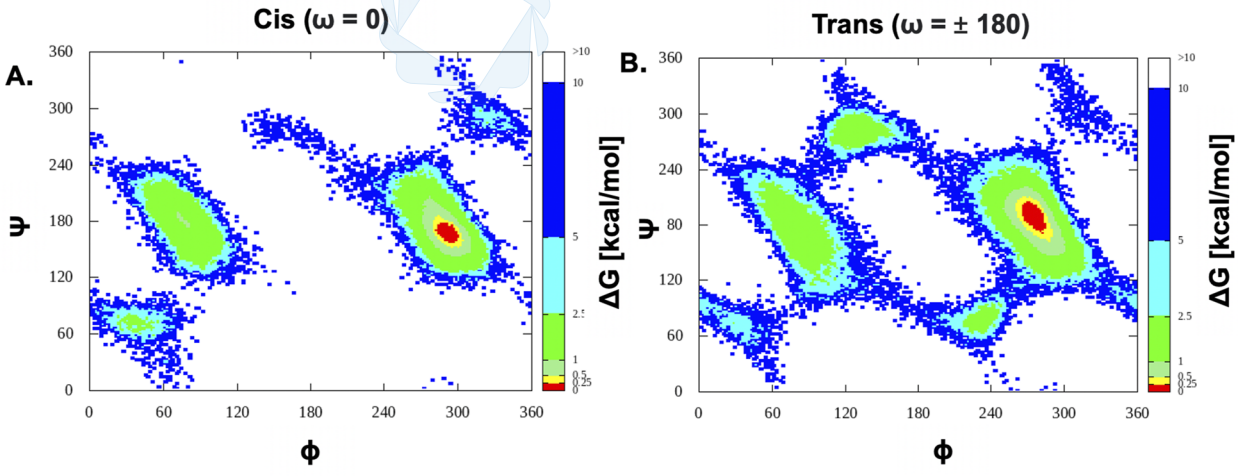

FIG. 2. Free energy Ramachandran plots for a selected residue in the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence. These plots are for a single peptoid in an aqueous solution. The cis (a) and trans (b) isomers are shown separately.

reported in experiments.[14,15] The CG simulations reproduce the helical conformation, among other structural properties. Furthermore, CG simulations of multiple peptoids yield a hemispherical aggregate that is in qualitative agreement with the microspheres reported in experiments.[14,15] The Nspe residues reside in the bulk of the hydrophobic core of the hemispherical aggregate, whereas the NLys residues are organized at the exterior of the aggregate. Finally, the Npmb residues are located along the curved interface of the aggregate.

Figures 2, SI.8, and SI.9 show that the AA simulations sample the helical conformations of the [Nspe–Npmb–$(Nspe)_2$–NLys–$Nspe]_2$ peptoid sequence. Figure 2 shows the Ramachandran plots for a single peptoid in an aqueous solution. The regions of energy minima correspond to helical conformations.[17,21,23,54] Similar helical conformations are observed in systems with multiple peptoids (see Figs. SI.8 and SI.9). It is noted that the Ramachandran plots show a preference toward either cis or trans isomers in multi-peptoid systems. In these systems, both isomers are rarely observed in the same peptoid residue. This observation is in agreement with other studies that report a bias toward a specific isomer in peptoid-based assemblies.[11,66]

The CG force field is developed by sampling the AA simulation trajectories. The force field development workflow is similar to an earlier study on the CG models for peptides.[20] The order of building the CG potentials follows a systematic routine described in a previous study.[19] The first set of steps is to derive the water–water and peptoid–peptoid potentials and the initial estimates for the bonded potentials. These steps can be performed in any order.

A. Water–water

An AA system with 2180 water molecules serves as the reference to develop CG water–water nonbonded potentials. The TIP3P model[67–69] is employed for the AA representation of the water molecules. The system is simulated in a box of dimensions 4 nm. The AA MD simulation is run for 200 ns in the NPT ensemble. The final 50 ns of the trajectory is sampled to generate a water–water RDF. This RDF is converted to a CG water–water potential using BI. The equilibrated AA configuration is converted to a CG configuration using the csg_map tool in the VOTCA package.[65] The CG simulation is run in the NVT ensemble using a 2 fs time step. The resulting water–water RDF from the CG trajectory does not agree with the corresponding RDF from the AA trajectory. To improve the CG potential, the IBI method is applied. After 300 steps of IBI, the water–water RDF from the CG trajectory is in complete agreement with the corresponding RDF for the AA trajectory (see Fig. SI.10). Each IBI step runs for 300 ps.

Since the reference system does not include any effects from other molecules, such as peptoids and ions, the CG water–water potential is transferable to systems with different concentrations of peptoids. Figure SI.11 shows that the water–water RDFs from the CG trajectory are in good agreement with the corresponding AA RDFs for low and high concentrations of peptoids.

B. Peptoid–peptoid

There are two types of peptoid–peptoid nonbonded interactions: intermolecular and intramolecular interactions. The intermolecular interactions govern the self-assembly of the peptoids, whereas the intramolecular interactions partially preserve the structure of individual peptoids. The FM with exclusions method,[20,42] in particular, resolves the intermolecular interactions between peptoids. The original implementation of the FM method[40] projects all the AA forces onto the CG coordinates. This would include effects of the bonded and intramolecular nonbonded interactions in the peptoid–peptoid nonbonded potentials. Hence, these interactions are excluded from the FM process. The peptoid–ion and ion–ion nonbonded interactions are developed in conjunction with the peptoid–peptoid potentials.

The intramolecular nonbonded interactions in peptoids are modeled with the same peptoid–peptoid potentials. Due to the exclusions, these CG potentials do not capture the underlying AA intramolecular effects that govern peptoid structure. This leads to a loss in the backbone conformation of the peptoids in the CG simulations. This is fixed with special bonded potentials that will be discussed in Sec. III E.

An AA system with a high concentration of peptoids is selected for the development of the peptoid–peptoid nonbonded potentials. The self-assembly of eight peptoids in the AA representation is modeled in a simulation box of dimensions 4 nm. The system is solvated with 2196 water molecules. Sixteen monovalent chloride ions are added to maintain charge neutrality. This system corresponds to a 0.65M concentration of peptoids. The AA simulation is run for 1000 ns in the NPT ensemble (Fig. SI.12). Since this AA system partially captures the aggregation of peptoids, it serves as a good reference for simulating the aggregation in the CG model. The forces are extracted from the final 200 ns of the simulation trajectory. The FM method compiles the forces to generate the peptoid–peptoid nonbonded potentials. These potentials are best suited for CG systems with the same concentration of peptoids as the AA reference system, i.e., 0.65M. The peptoid–peptoid nonbonded potentials govern the individual structure and assembly of the peptoids in the CG simulations.

C. Peptoid–water

The IBI method is used to generate peptoid–water potentials that resolve the solvation of the peptoid molecule. The peptoid–peptoid and water–water nonbonded potentials and the initial estimates of the bonded potentials are derived prior to this step. (The initial estimates for bonded potentials are derived using the BI method. See the section titled "Iterative refinement of bonded potentials" for details.) An AA simulation of a single peptoid in water is run for 1000 ns. The reference RDFs are constructed from the final 100 000 coordinate frames of the AA simulation trajectory. Hundred steps of IBI generate good quality peptoid–water potentials. Each IBI step executes a 300 ps CG MD simulation. The resulting CG potentials accurately resolve the overall solvation of the peptoid (see Fig. 3). Correct solvation of the peptoid is required to access the appropriate backbone conformations.[22,23] Minor errors are observed between the AA and CG peptoid–water distributions. Figure 3(d) shows that the backbone residues of the peptoid are slightly more hydrophilic in the CG representation. However, the overall description of the AA RDF is captured by the CG RDF. The ion–water potential is developed along with the peptoid–water potentials. Figure SI.13 shows that the AA and CG RDFs for ion–water interactions are in agreement with each other.

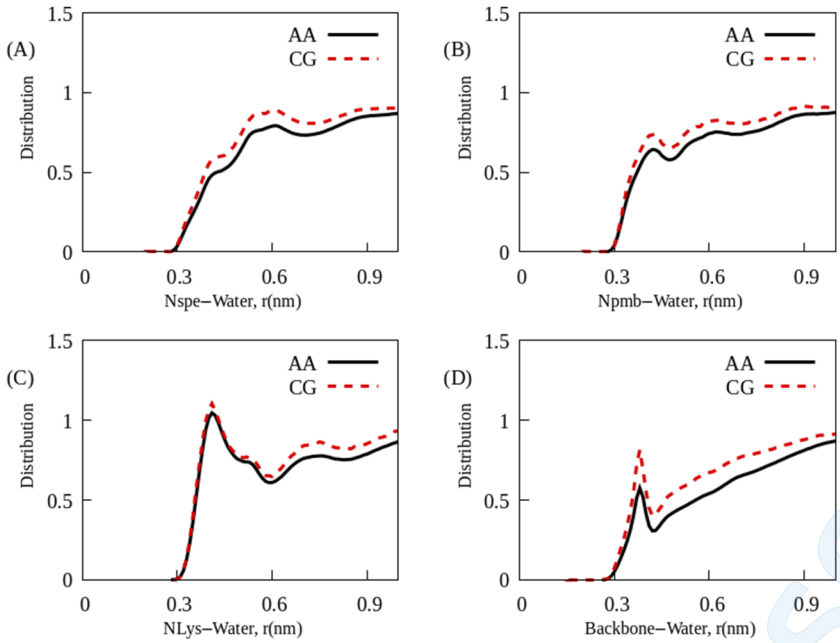

FIG. 3. Peptoid–water RDFs. (a)–(c) Solvation of the peptoid side chain residues. (d) Solvation of the peptoid backbone residues. The black and red distributions correspond to the measurements performed using the AA and CG trajectories, respectively.

D. Iterative refinement of bonded potentials

The BI method is employed to generate CG potentials for all bonded interactions, namely, bonds, angles, and dihedrals. An AA system encompassing the $[\text{Nspe–Npmb–(Nspe)}_2\text{–NLys–Nspe}]_2$ peptoid sequence in an aqueous solution is considered as the reference. This AA system is simulated for 1000 ns. The BI method processes the last 100 000 frames of the AA MD trajectory to generate the CG potentials. The resultant CG distributions are not in complete agreement with the corresponding AA distributions. Hence, the CG bonded potentials are iteratively refined using IBI. Figure 4 shows the resultant CG distributions to be in good agreement with the AA distributions.

The bonded potentials are refined in the order of increasing flexibility, i.e., bonds < angles < dihedrals.[19] Bonds and angles are refined with 3 and 2 steps of IBI, respectively. With these refinements, Fig. 4(c) shows that the CG dihedral distributions are in perfect agreement with the corresponding AA distributions. Hence, the dihedrals are not refined using IBI. The good agreement between the dihedral distributions demonstrates that the secondary structure of the peptoid, namely, the helical conformations, is preserved in the CG model.

As an aside, IBI is only performed on bonds and angles that involve backbone beads. Bonded interactions involving backbone–side chain and side chain–side chain interactions are well represented by the CG potentials generated via BI.

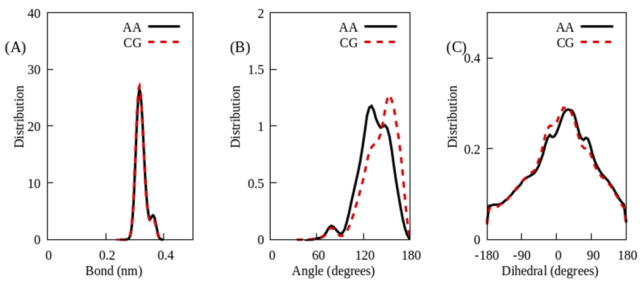

FIG. 4. Comparison of the bonded distributions across AA and CG simulations. (a) Average distribution of all bonds along the peptoid backbone. (b) and (c) Average distribution of all angles and dihedrals along the peptoid backbone, respectively. The black and red distributions correspond to the measurements performed using the AA and CG trajectories, respectively.

E. Peptoid chain conformations

All bonded and nonbonded CG potentials work in conjunction to govern the equilibrium chain conformation of the peptoid. To access the overall conformation of the peptoid, the end-to-end distance of the sequence is measured. This is the distance between the C and N termini of the peptoid sequence. Figure 5(a) shows that the end-to-end distance of the peptoid in the CG simulation is not in agreement with the corresponding distribution in the AA simulation. In particular, the CG representation of the peptoid samples smaller values of the end-to-end distance. These conformations could be due to the unrestrained flexibility of the peptoid backbone. It is surmised that the intramolecular nonbonded interactions play a role in maintaining the conformation of the peptoid backbone. Since these interactions are not sampled in the model development process, the CG backbone conformation is not in agreement with the corresponding results from the AA reference. For further assessment, the degree of bending of the peptoid backbone is measured with the 1–3–5 angles ($\text{BB}_X\text{–BB}_{X+2}\text{–BB}_{X+4}$).

Figure 5(b) shows that the CG distribution samples smaller angles between 60° and 120°. This indicates that the CG peptoid

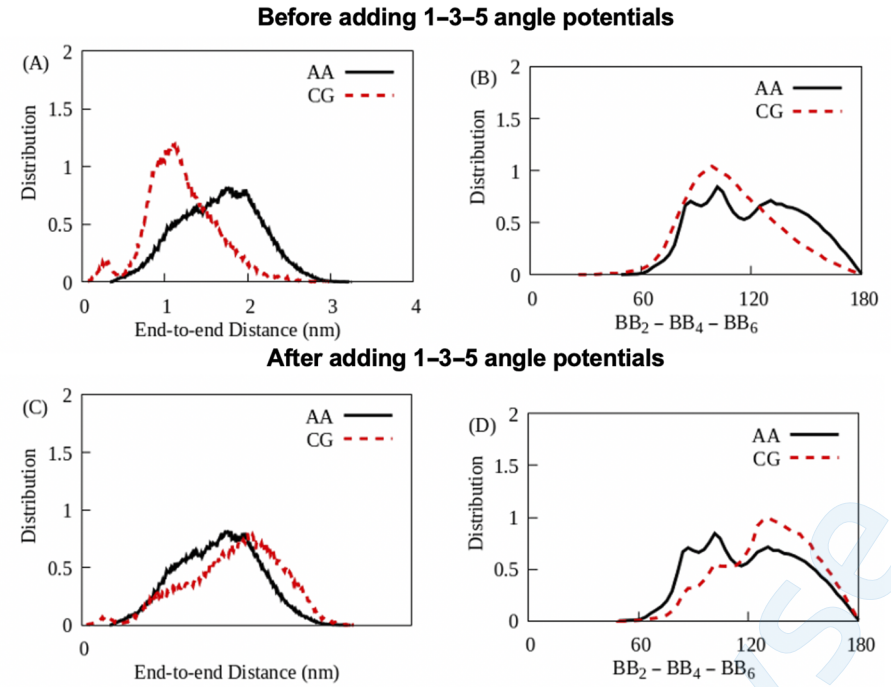

FIG. 5. End-to-end distance and a selected 1–3–5 angle of the peptoid measured using the AA and CG trajectories. (a) and (b) Distributions before adding 1–3–5 angle potentials to the CG force field. (c) and (d) Distributions after adding 1–3–5 angle potentials to the CG force field. The black and red distributions correspond to the measurements performed using the AA and CG trajectories, respectively.

backbone has a higher degree of bending in comparison with the AA reference. Hence, special 1–3–5 angles (BB_X–BB_{X+2}–BB_{X+4}) are added to the CG model. These angles are along the backbone of the peptoid sequence. Initially, these special potentials are derived by BI. Next, they are refined with a single step of IBI. These modifications result in the overall agreement of the end-to-end distance measured using the AA and CG trajectories [see Fig. 5(c)]. This means that the overall backbone conformation of the CG peptoid is in agreement with the AA reference. However, the 1–3–5 angle distribution [Fig. 5(d)] is still not in complete agreement with the corresponding AA distribution. A higher number of IBI steps do not improve the agreement between the AA and CG distributions of the 1–3–5 angles. Since the overall backbone conformation (end-to-end distance) is in agreement with the corresponding results from the AA simulations, the discrepancy in 1–3–5 angle distributions can be ignored. Figure 5(c) shows that the peak positions (at ~2 nm) of the AA and CG distributions are in agreement. In addition, it is significant that the CG simulation samples the entire AA distribution from ~0.5 to 3 nm. The minor disagreements in the distributions of the end-to-end distance could be due to the large breadth of the distribution. The agreement between the distributions can be improved by including additional CG special potentials, such as 1–5 bonds (BB_X–BB_{X+4}). This step is not performed to avoid over-constraining the peptoid backbone.

F. Comparison of time scales

Coarse-grained models accelerate the dynamics of peptoids due to a smoother potential energy surface.[70,71] Hence, the AA and CG time scales cannot be compared directly to each other. In analogy to an earlier study,[20] the diffusion coefficients of a single peptoid are compared across the AA and CG simulations. Table SI.I reports the self-diffusion coefficients of the peptoid in AA (D_{AA}) and CG (D_{CG}) representations. For a single peptoid in an aqueous solution, the ratio of D_{CG} to D_{AA} is ~4. This means that the effective time scale in CG simulations is four times than that for the AA simulations. However, this value is not applicable to systems with multiple peptoids (discussed in Sec. III G) as the acceleration in dynamics is dependent upon the concentration of peptoids. Hence, the AA and CG time scales are not compared for systems with multiple peptoids. In this case, the final equilibrated configuration in the CG simulation will be compared to the corresponding configuration in the AA simulation. Since the effective CG time scale is not computed for multiple peptoid simulations, the effective run-times for CG simulations (total number of MD time steps × actual time step × effective CG time scale) are not provided in Sec. III G. Instead, the total number of MD time steps is provided.

G. Systems with multiple peptoids

The CG model is tested on systems with multiple peptoids. The structure of individual peptoids and their assembly is validated against corresponding results from the AA simulations and experimental results. The results of CG systems encompassing a smaller number of peptoids (1, 2, 4, and 8 peptoids in water) are compared to the corresponding results from AA simulations. However, this comparison cannot be made for larger CG systems (namely, 64 and 128 peptoids in water). The AA simulations are unable to resolve the dynamics of a large number of peptoids across extensive spatiotemporal scales. This is attributed to the rough potential energy surface of the AA model that results in longer relaxation times. Hence, the results from the larger CG systems are compared with those from experiments.

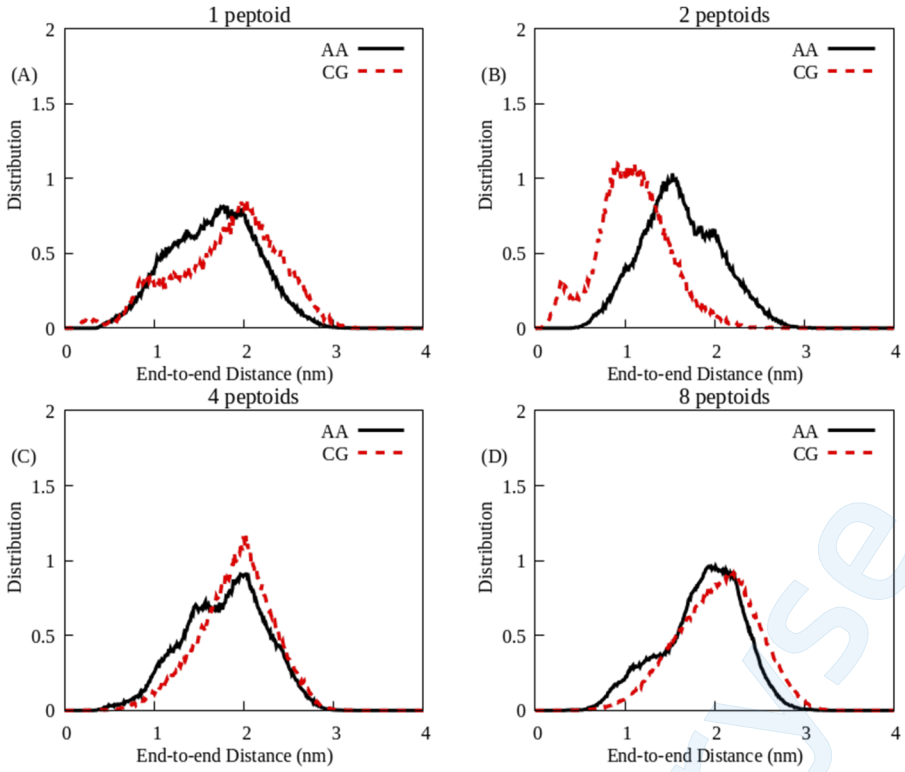

FIG. 6. End-to-end distance of the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence in systems with multiple peptoids. The black and red distributions correspond to the measurements performed using the AA and CG trajectories, respectively. Panel (a) is the same as Fig. 5(c).

For systems encompassing a smaller number of peptoids, the structure of the individual peptoids are compared across AA and CG simulations. Figure 6 shows a comparison of the end-to-end distances for 1, 2, 4, and 8 peptoids. With the exception of the 2 peptoid system, the measurements for the CG systems are in agreement with the corresponding results from the AA systems. This demonstrates that the CG model preserves the individual structure of the peptoids in the presence of interactions with neighboring peptoids. The effect of neighboring peptoid chains on the structure of individual peptoids was not used to parameterize the CG model and, hence, showcases the robustness and transferability of the model.

The tendency to retain helical conformations in CG simulations is assessed. The Ramachandran plots for the AA simulations of single and multiple peptoid systems show that the peptoids retain their helical conformations (Figs. 2, SI.8, and SI.9). Ramachandran plots cannot be constructed for the CG representation of the peptoids as it is only applicable for atomistic dihedrals. However, the distributions of the CG backbone dihedral (BB_X–BB_{X+1}–BB_{X+2}–BB_{X+3}) are indicative of the average backbone conformation. Hence, Fig. 7 shows a comparison of the distribution for the backbone dihedrals for 1, 2, 4, and 8 peptoid systems. With the exception of the 2 peptoid system, all distributions of the dihedral in the CG representation are in agreement with corresponding distributions in the AA representation. This demonstrates that the peptoids retain their helical conformations.

The discrepancy in the 2 peptoid system could be attributed to the bonded potentials that are fit to the single peptoid system. With the exception of the 2 peptoid system, the AA end-to-end distance for all systems is ~2 nm (see Fig. 6). The AA end-to-end distance in the 2 peptoid system is shorter than in the other systems (~1.5 nm). This means that the overall conformation of the peptoids in the 2 peptoid system is different from the overall conformation of peptoids in the other systems. It is surmised that this effect is overestimated in the CG 2 peptoid system wherein the end-to-end distance is even shorter (~1 nm). Furthermore, a shorter end-to-end distance in the AA 2 peptoid system affects the associated AA dihedral distribution (Fig. 7). Barring the 2 peptoid system, all dihedral distributions are unimodal in nature; i.e., they have a single peak centered at ~0°. However, the AA dihedral distribution for the 2 peptoid system is bimodal; i.e., there are peaks at −90° and 90°. The CG model is unable to capture this change in the backbone conformation. Taken together, the AA 2 peptoid simulations sample drastically different conformations in comparison with the other systems. The CG model provides a good agreement for the 4 and 8 peptoid systems as the corresponding AA distributions are not drastically different from the reference system (namely, the AA single peptoid system that has an average end-to-end distance of 2 nm, and the peak of the dihedral distribution is centered at ~0°). Figure SI.14 shows the dimer formed by the 2 AA peptoids at different instances of time in the AA MD trajectory. Figures SI.14A and SI.14B show configurations where the 2 peptoids adopt a shorter end-to-end distance, whereas Figs. SI.14C and SI.14D show configurations where the 2 peptoids adopt relatively longer end-to-end distances. Since the average end-to-end distance is short for these AA peptoids (1.5 nm), it is surmised that the conformations shown in Figs. SI.14A and SI.14B are more likely. The reasons for adopting shorter conformations could be attributed

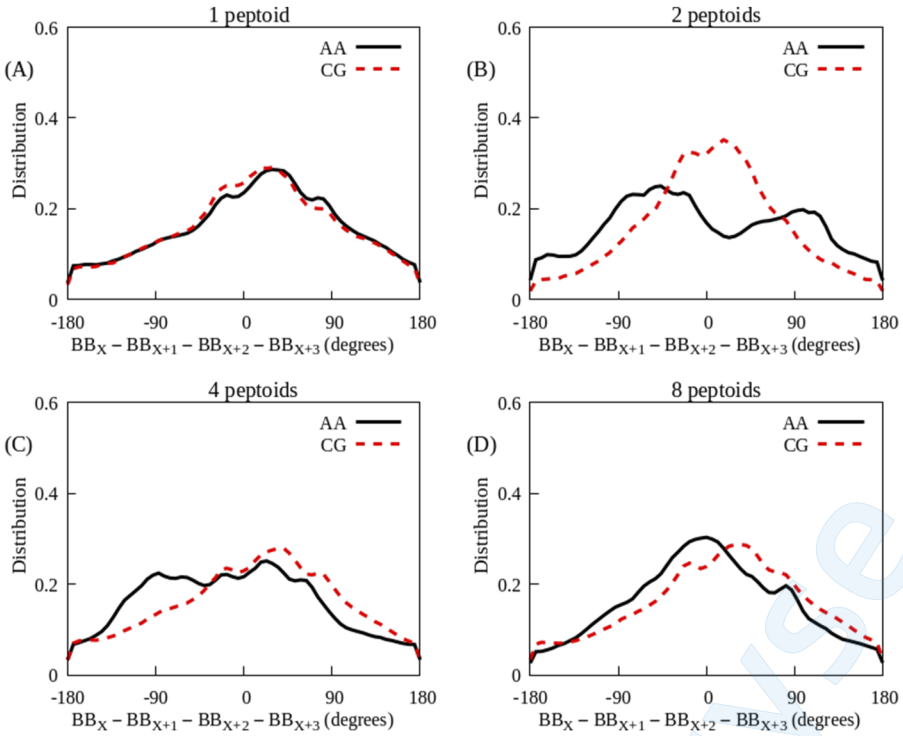

FIG. 7. Comparison of the backbone dihedral distributions between AA and CG representations of the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence. The black and red distributions correspond to measurements performed using the AA and CG trajectories, respectively.

to the hydrophobic effect. However, this would require a more rigorous investigation of how the peptoid residues pack in the dimer. Since the goal was to assess the transferability of the CG model, further investigations related to the 2 peptoid system are beyond the scope of this work and will be pursued in the future.

In systems with a larger number of peptoids, the final equilibrated configuration in the CG simulation is compared with the microspheres reported by experimental studies.[14,15] Figure 8 shows that the 64 CG peptoids self-assemble to form a hemispherical aggregate. The curvature associated with this aggregate could explain the formation of large-scale curved assemblies, such as microspheres. Due to longer relaxation times in the AA model, peptoids do not assemble into aggregates that can be compared with experimental observations. For example, the AA system with 64 peptoids forms a loosely connected network of small aggregates (see Fig. SI.15).

Experiments report microspheres that have an average diameter of 3 μm[15] Since this dimension is much larger than the length of a single peptoid (~2 nm), experiments suggest that a microsphere may consist of a collection of smaller assemblies. In particular, experiments report that long peptoid aggregates self-assemble to form microspheres.[15] Figure SI.16 shows that 128 CG peptoids self-assemble into a long sheet-like aggregate. This aggregate encompasses two smaller hemispherical aggregates that are connected to each other. It is surmised that larger systems (with more than 128 peptoids) will form a relatively longer aggregate that will be a network of multiple hemispherical assemblies. However, with the current approach, the self-assembly of a larger number of peptoids is not computationally feasible. Instead, one can start with a preassembled initial configuration where the peptoids are placed in close proximity to each other. A good initial configuration will ensure that the final assembly is in agreement with experiments. The observations from this work can guide the design of a preassembled initial configuration for a larger system. For example, multiple hemispherical aggregates could be placed in close proximity and a long MD simulation could resolve the assembly of a larger structure that is in better agreement with experiments (microspheres). This approach can be pursed in the future.

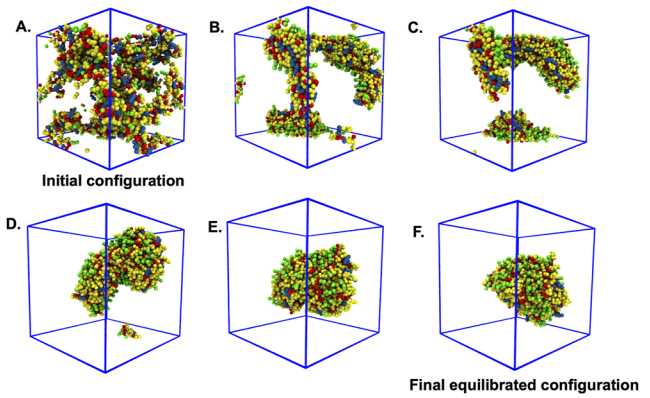

FIG. 8. Self-assembly of 64 CG peptoids into a hemispherical aggregate. (a)–(f) Instances of the self-assembly process at 0, 3 × 10^6, 6 × 10^6, 1 × 10^7, 1.5 × 10^7, and 2 × 10^7 time steps, respectively. Color scheme: red—peptoid backbone, yellow—Nspe, blue—Npmb, and green—NLys.

Although the final aggregate formed in the CG simulations cannot be compared with AA simulations, Fig. SI.17 shows that the local properties such as the Nspe–Nspe interactions are in qualitative agreement with the corresponding results from AA simulations. The location of the first peak (approximately between 0.3 and 0.5 nm) in the CG distributions is in agreement with the corresponding results from the AA distributions. However, the peak heights do not agree due to differences in the degree of aggregation in the AA and CG simulations. Previous studies suggest that aromatic stacking between Nspe residues could yield assemblies such as microspheres.[15] Furthermore, experiments on the [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$ peptoid sequence report that Nspe residues populate two faces of the peptoid helix.[14,15] The third face of the helix consists of alternating Npmb and NLys residues.[14,15] It is noted that the Npmb–Npmb and NLys–NLys interactions are not in agreement across the AA and CG simulations due to the differences in the aggregation of the peptoids. Hence, the organization of the residues on the third face of the helix is key to the difference in the results from the AA and CG simulations.

Experiments show that the residues on the third face of the peptoid helix determine the shape and size of the final assembly.[15] To understand the organization of the residues on the third face of the helix, the spatial location of Npmb and NLys residues in the CG system encompassing 64 peptoids is assessed (see Fig. 9). Figure 9(d) shows that NLys residues are located at the exterior of the final assembly. This is attributed to the hydrophilicity of these residues. It is noted that the final assembly has two exposed surfaces, namely, a curved and a flat surface (Fig. 10). There are a higher number of NLys residues on the curved surface. This is explained by the conformation adopted by the individual peptoids within the assembly. Peptoid conformation is assessed on the basis of the placement of the two NLys residues in the sequence. First, Fig. 10(b) depicts a peptoid conformation wherein there is one NLys residue on both of the exposed surfaces. Next, Fig. 10(b) depicts a peptoid conformation wherein both NLys residues are placed on the curved surface. The second conformation results in a higher number of NLys residues on the curved surface. The curvature of the surface allows the NLys residues to be farther apart from each other. This is expected as NLys residues electrostatically repel each other (due to their positive charges). This observation suggests that the conformation wherein both NLys residues are placed on the curved surface could be more favorable.

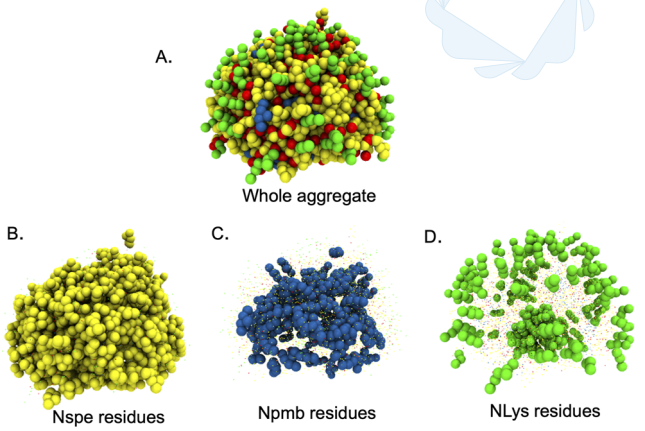

FIG. 9. (a) Hemispherical aggregate consisting of 64 CG peptoids. (b)–(d) Spatial location of the Nspe, Npmb, and NLys residues, respectively. Color scheme: red—peptoid backbone, yellow—Nspe, blue—Npmb, and green—NLys.

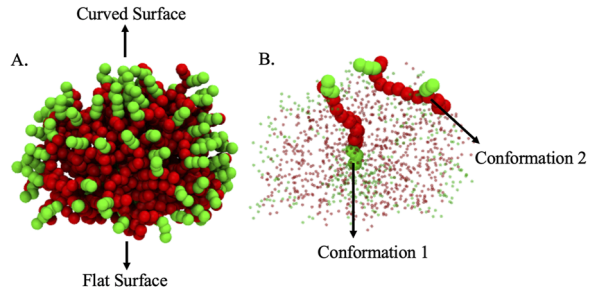

FIG. 10. (a) Hemispherical aggregate formed by 64 CG peptoids. The backbone (red) and NLys (green) beads are shown for an easier visualization. The curved and flat exterior surfaces are labeled. (b) The two conformations that the peptoids adopt in the hemispherical aggregate. Conformation 1: one NLys residue is located on the flat surface and one NLys residue is located on the curved surface. Conformation 2: both NLys residues are located on the curved surface. All other peptoids are represented as small spheres for a better visualization of the two conformations.

To identify the underlying cause for curvature in the final assembly, the population of the residues at the curved interface is investigated. This interface separates the hydrophilic exterior (primarily consisting of NLys residues) from the hydrophobic interior (primarily consisting of Nspe residues). Figures 9 and SI.18 shows that Npmb residues are preferentially located along the curved interface of the aggregate. Due to their mildly hydrophilic nature, these residues are placed along the interface of the aggregate.

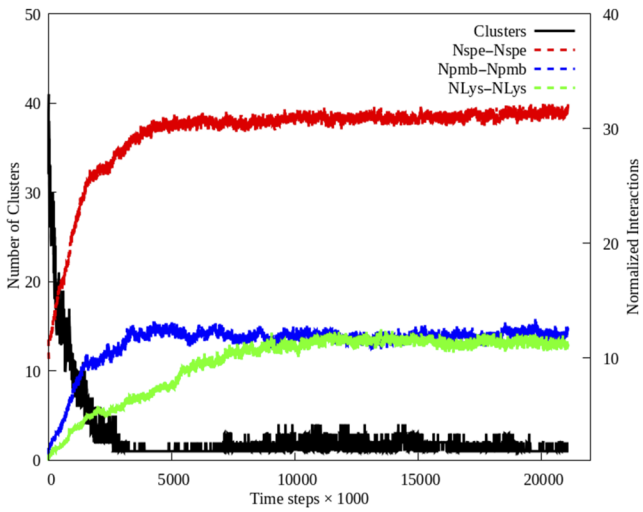

FIG. 11. Cluster count and interaction count measurements during the self-assembly of 64 CG peptoids in an aqueous solution. The black line represents the number of peptoid clusters (y-axis on the left) as a function of the number of time steps. The remaining lines represent the interactions between peptoid residues (y-axis on the right) as a function of the number of time steps.

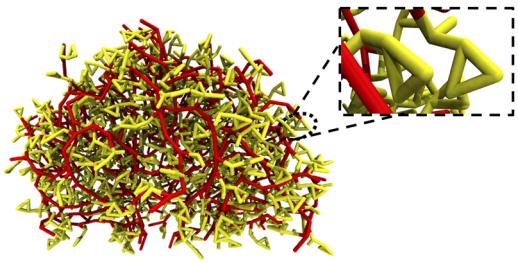

FIG. 12. Zoomed-in view of a possible instance of aromatic stacking. The hemispherical aggregate consisting of 64 peptoids is shown in a stick figure representation. The red and yellow sticks represent the backbone and Nspe residues, respectively. The other peptoid residues are not shown for clarity. The triangles represent the aromatic rings of Nspe residues.

Furthermore, it is surmised that the curvature of the aggregate enhances the exposure of these residues to water, which would be consistent with their mildly hydrophilic nature.

Finally, it is observed that the interactions between Nspe residues due to the hydrophobic effect drive the aggregation of the peptoids. Figure 11 shows that Nspe–Nspe, Npmb–Npmb and NLys–NLys interactions increase during the progression of the self-assembly process. The highest increase is in the Nspe–Nspe interactions. The steep increase in the Nspe–Nspe interactions is correlated with the steep decrease in the number of peptoid clusters. Hence, the Nspe residues play a critical role in the self-assembly of peptoids. Since these residues consist of chiral, aromatic side chains, it is surmised that the Nspe–Nspe interactions are indicative of the aromatic stacking between the peptoids (see Fig. 12). This local organization between the peptoid chains could stabilize the formation of large scale assemblies.[15]

IV. CONCLUSIONS

Protein mimics such as peptoids can form a variety of self-assembled nanostructures depending upon their sequence and secondary structures. The exact conformation and organization of these molecules within a nanostructure remains unknown. Of specific interest is a peptoid sequence[14,15] with a helical secondary structure that has been observed to assemble into microspheres. In this study, the conformation and organization of the peptoid sequence within an assembly is elucidated via the use of a hybrid, bottom-up coarse-graining approach. The peptoid sequence is [Nspe–Npmb–$(Nspe)_2$–NLys–Nspe]$_2$. The resultant CG model resolves the structure, solvation, and assembly of randomly dispersed peptoids in an aqueous solution into aggregates. The bottom-up coarse-graining approach incorporates the AA details into the CG model, thereby preserving the structure and conformation of the sequence.

The structural features, such as the helical conformation, overall backbone conformation, and solvation of the peptoid, are reproduced in the CG model. Furthermore, the model can efficiently resolve the aggregation of 64 peptoids into a hemispherical assembly that is in qualitative agreement with experiments. To better understand the internal structure of this assembly, the organization of the three peptoid residues (Npmb, NLys, and Nspe) is investigated. The Npmb residues are located along the curved interface of the hemispherical assembly, thereby governing the curvature of the aggregate. Hence, to tune the shape and size of microspheres, one can selectively mutate the Npmb residues. Furthermore, the organization of the peptoids within the assembly is characterized by two peptoid conformations. These conformations determine the distribution of the NLys residues on the exterior of the assembly. In the case of the hemispherical assembly, there are two exterior surfaces. One surface is curved and consists of a larger number of NLys residues. The other surface is flat and consists of a fewer number of NLys residues. Finally, the interactions between the Nspe residues due to the hydrophobic effect govern the self-assembly of the peptoids. The Nspe residues form the hydrophobic core of the aggregate.

This study shows that large peptoid-based assemblies could consist of a collection of smaller hemispherical aggregates. The 128 peptoid CG simulation shows that two hemispherical aggregates are connected to each other. This means that large peptoid-based assemblies (possibly microspheres that are observed in experiments) could consist of a network of multiple hemispherical aggregates. This observation could help in better understanding the material properties associated with peptoid-based assemblies, such as porosity and overall stability.

Fundamental insights into the organization of molecules and their conformations within assemblies could elucidate the impact of residue sequence on the structure, morphology, and function of the self-assembled nanostructures. This understanding will significantly accelerate the development of sequence–structure–property relations of synthetic and natural materials for their use in numerous disciplines, including pharmaceuticals,[6,72] biomedicine,[45–47] and electronics.[48,49] The hybrid coarse-graining approach adopted in this study can be used to support these efforts.

SUPPLEMENTARY MATERIAL

See the supplementary material for a description of all tools in a GitHub repository, simulation details, CG schemes of side chain residues, AA dihedral schematics, ω dihedrals plots, energies of ω dihedrals, Ψ–ω and Φ–ω plots, χ_1 dihedral plots, dihedral plots of a 8 AA peptoid system, water–water RDFs, self-assembly trajectory of an 8 AA peptoid system, ion–water RDF, comparison of AA and CG time scales for a single peptoid system, 2 AA peptoid dimer configurations, self-assembly of a 64 AA peptoid system, snapshots of the final assembly formed by the 128 CG peptoid system, Nspe–Nspe RDFs, and snapshots of the final assembly formed by the 64 CG peptoid system.

ACKNOWLEDGMENTS

M.D. would like to acknowledge NSF CAREER Grant No. DMR-1654325 and NSF Grant No. OAC-1835449 for financial support. The authors acknowledge NSF ACCESS (allocation Grant No. DMR-140125) for providing the computational resources used in this study. A.B. would like to thank Dr. Christoph Junghans from Los Alamos National Laboratory for helpful discussions regarding the usage of the VOTCA package.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Akash Banerjee: Formal analysis (lead); Investigation (lead); Methodology (supporting); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Meenakshi Dutt**: Conceptualization (lead); Funding acquisition (lead); Methodology (lead); Project administration (lead); Supervision (lead); Writing – original draft (supporting); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in the GitHub repository Multiscale-Multiresolution-CG-Models-for-Peptoids at Ref. 52.

REFERENCES

[1] J. Laxio Arenas, J. Kaffy, and S. Ongeri, Curr. Opin. Chem. Biol. **52**, 157 (2019).
[2] K. D. Stigers, M. J. Soth, and J. S. Nowick, Curr. Opin. Chem. Biol. **3**, 714 (1999).
[3] B. M. deRonde and G. N. Tew, Pept. Sci. **104**, 265 (2015).
[4] N. T. Ross, W. P. Katt, and A. D. Hamilton, Philos. Trans. R. Soc. London, Ser. A **368**, 989 (2010).
[5] R. N. Zuckermann, J. M. Kerr, S. B. H. Kent, and W. H. Moos, J. Am. Chem. Soc. **114**, 10646 (1992).
[6] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, and C. K. Marlowe, Proc. Natl. Acad. Sci. U. S. A. **89**, 9367 (1992).
[7] S. M. Miller, R. J. Simon, S. Ng, R. N. Zuckermann, J. M. Kerr, and W. H. Moos, Drug Dev. Res. **35**, 20 (1995).
[8] J. Sun and R. N. Zuckermann, ACS Nano **7**, 4715 (2013).
[9] H. K. Murnen, A. M. Rosales, J. N. Jaworski, R. A. Segalman, and R. N. Zuckermann, J. Am. Chem. Soc. **132**, 16112 (2010).
[10] E. J. Robertson, G. K. Olivier, M. Qian, C. Proulx, R. N. Zuckermann, and G. L. Richmond, Proc. Natl. Acad. Sci. U. S. A. **111**, 13284 (2014).
[11] R. V. Mannige, T. K. Haxton, C. Proulx, E. J. Robertson, A. Battigelli, G. L. Butterfoss, R. N. Zuckermann, and S. Whitelam, Nature **526**, 415 (2015).
[12] H. Jin, Y.-H. Ding, M. Wang, Y. Song, Z. Liao, C. J. Newcomb, X. Wu, X.-Q. Tang, Z. Li, Y. Lin, F. Yan, T. Jian, P. Mu, and C.-L. Chen, Nat. Commun. **9**, 270 (2018).
[13] X. Jiang, R. K. Spencer, J. Sun, C. Ophus, R. N. Zuckermann, K. H. Downing, and N. P. Balsara, J. Phys. Chem. B **123**, 1195 (2019).
[14] G. R. Perez Bakovic, J. L. Roberts, B. Colford, M. Joyce, and S. L. Servoss, Biopolymers **110**, e23283 (2019).
[15] M. L. Hebert, D. S. Shah, P. Blake, J. P. Turner, and S. L. Servoss, Org. Biomol. Chem. **11**, 4459 (2013).
[16] S. Mukherjee, G. Zhou, C. Michel, and V. A. Voelz, J. Phys. Chem. B **119**, 15407 (2015).
[17] P. Armand, K. Kirshenbaum, A. Falicov, R. L. Dunbrack, K. A. Dill, R. N. Zuckermann, and F. E. Cohen, Folding Des. **2**, 369 (1997).
[18] C. Peter and K. Kremer, Soft Matter **5**, 4357 (2009).
[19] O. Bezkorovaynaya, A. Lukyanov, K. Kremer, and C. Peter, J. Comput. Chem. **33**, 937 (2012).
[20] A. Banerjee, C. Y. Lu, and M. Dutt, Phys. Chem. Chem. Phys. **24**, 1553 (2022).
[21] S. H. Park and I. Szleifer, J. Phys. Chem. B **115**, 10967 (2011).
[22] D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, and S. Whitelam, J. Comput. Chem. **35**, 360 (2014).
[23] L. J. Weiser and E. E. Santiso, J. Comput. Chem. **40**, 1946 (2019).
[24] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, Jr., J. Comput. Chem. **31**, 671 (2010).
[25] V. A. Voelz, K. A. Dill, and I. Chorny, Biopolymers **96**, 639 (2011).
[26] S. Hoyas, V. Lemaur, Q. Duez, F. Saintmont, E. Halin, J. De Winter, P. Gerbaux, and J. Cornil, Adv. Theory Simul. **1**, 1800089 (2018).
[27] S. Jiao, A. DeStefano, J. I. Monroe, M. Barry, N. Sherck, T. Casey, R. A. Segalman, S. Han, and M. S. Shell, Macromolecules **54**, 5011 (2021).
[28] M. Zhao, K. J. Lachowski, S. Zhang, S. Alamdari, J. Sampath, P. Mu, C. J. Mundy, J. Pfaendtner, J. J. De Yoreo, C.-L. Chen, L. D. Pozzo, and A. L. Ferguson, Biomacromolecules **23**, 992 (2022).
[29] J. R. Edison, R. K. Spencer, G. L. Butterfoss, B. C. Hudson, A. I. Hochbaum, A. K. Paravastu, R. N. Zuckermann, and S. Whitelam, Proc. Natl. Acad. Sci. U. S. A. **115**, 5647 (2018).
[30] W. G. Noid, J. Chem. Phys. **139**, 090901 (2013).
[31] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, J. Chem. Theory Comput. **4**, 819 (2008).
[32] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).
[33] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink, J. Chem. Theory Comput. **9**, 687 (2013).
[34] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt, Soft Matter **9**, 2108 (2013).
[35] M. Zhao, J. Sampath, S. Alamdari, G. Shen, C.-L. Chen, C. J. Mundy, J. Pfaendtner, and A. L. Ferguson, J. Phys. Chem. B **124**, 7745 (2020).
[36] A. Villa, N. F. A. van der Vegt, and C. Peter, Phys. Chem. Chem. Phys. **11**, 2068 (2009).
[37] T. K. Haxton, R. V. Mannige, R. N. Zuckermann, and S. Whitelam, J. Chem. Theory Comput. **11**, 303 (2015).
[38] D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[39] F. Müller-Plathe, ChemPhysChem **3**, 754 (2002).
[40] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[41] A. Villa, C. Peter, and N. F. A. van der Vegt, Phys. Chem. Chem. Phys. **11**, 2077 (2009).
[42] V. Rühle and C. Junghans, Macromol. Theory Simul. **20**, 472 (2011).
[43] R. D. Hills, Jr., L. Lu, and G. A. Voth, PLoS Comput. Biol. **6**, e1000827 (2010).
[44] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophys. J. **92**, 4289 (2007).
[45] Z. S. Clauss and J. R. Kramer, ACS Appl. Mater. Interfaces **14**, 22781 (2022).
[46] B. Zhang, M. Li, M. Lin, X. Yang, and J. Sun, Biomater. Sci. **8**, 6969 (2020).
[47] J. Sun and Z. Li, *Peptide Applications in Biomedicine, Biotechnology and Bioengineering* (Woodhead Publishing, 2018), p. 183.
[48] J. Sun, X. Jiang, A. Siegmund, M. D. Connolly, K. H. Downing, N. P. Balsara, and R. N. Zuckermann, Macromolecules **49**, 3083 (2016).
[49] B. Kang, W. Yang, S. Lee, S. Mukherjee, J. Forstater, H. Kim, B. Goh, T.-Y. Kim, V. A. Voelz, Y. Pang, and J. Seo, Sci. Rep. **7**, 4786 (2017).
[50] J. Sun, G. M. Stone, N. P. Balsara, and R. N. Zuckermann, Macromolecules **45**, 5151 (2012).
[51] J. Sun, X. Liao, A. M. Minor, N. P. Balsara, and R. N. Zuckermann, J. Am. Chem. Soc. **136**, 14990 (2014).
[52] A. Banerjee and M. Dutt, Multiscale Multiresolution CG Models for Peptoids, https://github.com/duttm/Multiscale-Multiresolution-CG-Models-for-Peptoids, 2022; accessed January 2023.
[53] P. Gao and A. Tartakovsky, arXiv:1903.01975 (2019).
[54] G. L. Butterfoss, P. D. Renfrew, B. Kuhlman, K. Kirshenbaum, and R. Bonneau, J. Am. Chem. Soc. **131**, 16798 (2009).
[55] J. W. Laura and E. S. Erik, AIMS Mater. Sci. **4**, 1029 (2017).
[56] M. Yasodharababu, S. L. Servoss, and A. K. Nair, J. Biomech. **121**, 110381 (2021).
[57] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. Mackerell, Jr., J. Chem. Theory Comput. **8**, 3257 (2012).
[58] A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).
[59] Q. Sui, D. Borchardt, and D. L. Rabenstein, J. Am. Chem. Soc. **129**, 12042 (2007).

[60]M. Feigel, J. Phys. Chem. **87**, 3054 (1983).
[61]A. Prakash, C. D. Fu, M. Bonomi, and J. Pfaendtner, J. Chem. Theory Comput. **14**, 4985 (2018).
[62]J. Pfaendtner and M. Bonomi, J. Chem. Theory Comput. **11**, 5062 (2015).
[63]A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).
[64]F. P. Gasparro and N. H. Kolodny, J. Chem. Educ. **54**, 258 (1977).
[65]V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009).
[66]D. R. Greer, M. A. Stolberg, J. Kundu, R. K. Spencer, T. Pascal, D. Prendergast, N. P. Balsara, and R. N. Zuckermann, J. Am. Chem. Soc. **140**, 827 (2018).
[67]P. Mark and L. Nilsson, J. Phys. Chem. A **105**, 9954 (2001).
[68]E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[69]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[70]D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer, Phys. Chem. Chem. Phys. **13**, 10412 (2011).
[71]S. J. Marrink and D. P. Tieleman, Chem. Soc. Rev. **42**, 6801 (2013).
[72]G. Diamond, N. Molchanova, C. Herlan, J. A. Fortkort, J. S. Lin, E. Figgins, N. Bopp, L. K. Ryan, D. Chung, R. S. Adcock, M. Sherman, and A. E. Barron, Pharmaceuticals **14**, 304 (2021).