JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

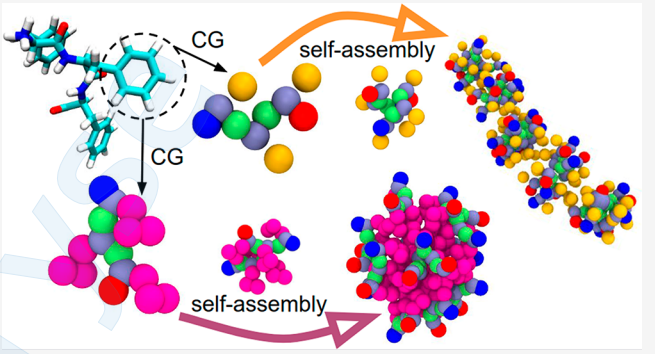
Multiscale, Multiresolution Coarse-Grained Model via a Hybrid Approach: Solvation, Structure, and Self-Assembly of Aromatic Tripeptides

Mason Hooten, Akash Banerjee, and Meenakshi Dutt*

Cite This: https://doi.org/10.1021/acs.jctc.3c00458 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Short aromatic peptides have been observed to assemble into diverse nanostructures, including fibers, tubes, and vesicles, using computational techniques. However, the computational studies have employed top-down coarse-grained (CG) models, which are unable to capture the assembly along with the conformation, packing, and organization of the peptides within the aggregates in a manner that is consistent with the all atom (AA) representation of the molecules. In this study, a hybrid structure- and force-based approach is adapted to develop a bottom-up CG force field of triphenylalanine using reference data from AA trajectories. This approach follows a flexible methodology to approximate the chemical complexity of the underlying AA representation with the chosen CG representation. Two CG models are developed with distinct representations of the aromatic side chains. The first uses a simple single-bead representation, while the second uses a three-bead representation to more accurately represent the planarity of the ring. The one-bead model yields nanorods, while the three-bead model results in nanospheres. The role of different chemical groups in the assembly of nanostructures is identified, along with the importance of steric effects on the packing of the peptides within assemblies.



INTRODUCTION

The assembly of aromatic peptides is pertinent to diverse disciplines, including human health and electronics. In the overarching area of human health, the assembly of aromatic peptides has been attributed to the kinetics of formation and thermodynamics of the stability of amyloid proteins.[1,2] These proteins have been found as deposits in several diseases including Alzheimer's Disease and Parkinson's Disease.[3] Another example is self-assembled nanostructures, such as vesicles, encompassing aromatic peptides that have been used to encapsulate, store, and deliver therapeutics to targeted regions in the human body.[4−7] In the broad area of electronics, self-assembly of short aromatic peptides has been observed to yield stable nanotubes[8] which have been further utilized to grow metallic and semiconducting nanowires. Another example is the self-assembly of π-conjugated oligopeptides with aromatic cores that have yielded nanostructures with optoelectronic properties.[9] The diverse functionalities of the nanostructures encompassing aromatic peptides are a testament to their multifunctionality and are predicated upon the packing and organization of the aromatic groups within the self-assembled nanostructures. The latter is dependent upon the conformation of the peptides within the assemblies. Given the wide prevalence of self-assembled peptides with aromatic residues in diverse disciplines of critical importance to society, it becomes imperative to understand the packing and organization of the molecules within assemblies encompassing aromatic peptides. This study focuses on the self-assembly of the small aromatic tripeptide Triphenylalanine (FFF) along with their molecular conformation, packing, and organization within the assembly. This peptide sequence is chosen due to its simplicity.

Experimental studies[10] on the assembly of charged-terminal-capped FFF peptides (N-terminal with NH_3^+, C-terminal with COO^-) reported the formation of plate-like structures with β-sheet content. Another study[11] using uncapped FFF peptides reported diverse self-assembled structures including laminated helical-ribbons, leaflike dendrimers, flowers, doughnuts, and needles. The results seemed to indicate that the FFF molecules adopt both parallel and antiparallel β-sheets, although the former is surmised to be energetically preferred due to the $\pi-\pi$ stacking between the aromatic rings of molecules that have

Special Issue: Computational and Theoretical Studies Focused on Self-Assembly and Molecular Organization

Received: April 30, 2023
Revised: October 13, 2023
Accepted: October 16, 2023

hydrogen bonding. Still another study reported nanofiber aggregates encompassing uncapped FFF.[12]

Using different capping groups[13] such as *tert*-butyloxycarbonyl (Boc) group for the N terminal and carboxylic acid for the C terminal, the FFF derivative adopts a turnpike conformation and self-assembles into crystalline nanospheres which are thermally and chemical stable. FFF capped with two fluorenyl functionalities self-assembled into stacked braids which further formed corkscrew-like microstructures at higher peptide concentrations.[14] Microtubes were also observed under different solvent conditions. Peptides capped with fluorenyl and benzyl groups have formed spherulitic or sheaf-like aggregates, as well as flower-like aggregates comprising twisted ribbons with a common spherical origin.[15] Hybridized L- and D-phenylalanine peptides have produced γ-turn oligomers with high thermal stability.[16] FFF has also been grafted to synthetic polymers to induce the self-assembly of fiber networks to form bioinspired hydrogels.[17] However, due to constraints in resolution, experimental approaches are unable to provide insight into the mechanisms and processes underlying the self-assembly of FFF. This difficulty can be addressed via the adoption of suitable computational approaches.

All atom (AA) Molecular Dynamics (MD) simulations[10] have been used to investigate the assemblies formed by FFF peptides. Due to the computational cost of running AA MD simulations, 8 FFF peptides were simulated using the generalized Born with simple smoothing (GBSW) implicit solvent model. FFF was observed to assemble into open and closed networks. Also, the two-stranded sheets always had antiparallel organization. However, the three-stranded sheets had both parallel and antiparallel organization. Whereas AA MD simulations provide insight into the detailed dynamics of the molecules while accurately representing the chemistry of the system, they are computationally expensive. Beginning from a random dispersion of molecules in solution, their self-assembly spans multiple length and time scales. Hence, the process of self-assembly is inherently multiscale. The extended dynamics underlying the self-assembly process can be captured using reduced models, such as coarse-grained (CG) representations of the molecules.[18−23] To that end, another study[24] adopted the CG explicit solvent Martini model to resolve the mechanisms underlying the self-assembly of FFF. The study reported the spontaneous formation of solid core nanospheres and nanorods within which the peptides were predominantly antiparallel to each other. This enabled the aggregates encompassing FFF to form β-sheet-like structures. The study used a top-down CG model[20,25] which successfully resolved the morphology of the intermediate aggregates and equilibrium nanostructures resulting from the self-assembly of FFF peptides. However, the study did not elucidate the conformation, packing, or organization of these molecules within the aggregates or nanostructures.

Bottom-up coarse-grained models are required to capture the conformation, packing, and organization of peptides within aggregates.[26] These CG models are parametrized with data derived from AA MD simulations. This approach emphasizes the influence of the local structure and forces that govern the conformation of individual peptides, as observed in the AA reference. Previous studies have employed various techniques to develop bottom-up CG models.[26−31]

Nonbonded potentials play an especially important role in CG models, where the strength of the solute interactions will largely determine the association and packing of many solute molecules. All CG potentials attempt to reproduce the effective potential energy surface of the reference system, usually represented as a potential of mean force (PMF).[28] Villa and colleagues present models in which nonbonded potentials are derived by estimating the PMF from simulations of a pair of solute molecules whose pair distance is constrained.[30,31] In the model with explicit solvent effects,[31] multibody effects are eliminated from CG pair potentials by iteratively targeting the AA PMF. Carmichael and Shell apply relative entropy minimization, which aims to minimize the difference between the configurational ensembles of modeled and reference systems.[32] Ozgur and Sayar parameterize a reduced set of nonbonded interactions with functional forms designed to directly reproduce the secondary structures observed in relevant experiments.[33]

This study adapts structure- and force-based methods to parameterize two CG models to approximate the physics of an underlying AA model. This hybrid approach has been previously implemented to study systems with multiple amphiphilic peptides in aqueous solution.[20,34] A structure-based method, i.e., the Iterative Boltzmann Inversion (IBI) technique, is used to sample the bonded interactions in the peptide. This method samples the bonded distributions in the AA MD trajectories to derive the CG potentials. IBI ensures that the structure of the individual peptides is preserved in the CG simulations. However, it has been shown that IBI does not work well for nonbonded interactions, namely, interpeptide interactions. IBI generates multiwell potentials that are unsuitable for the transferability of the CG force field to other peptide concentrations. Hence, force-matching (FM), i.e., a force-based method, is employed to derive the nonbonded potentials. FM generates single-well potentials that can be transferred within a limited peptide concentration range. Finally, the interactions between the peptides and water are derived by IBI. This ensures that the structure of the peptide−water interface is preserved in the CG simulations. This hybrid approach has been adapted to capture the solvation, structure, and self-assembly of lipid-like aliphatic peptides[20] and amphiphilic helical peptoids[18] along with their conformation, packing, and organization within the aggregates.

In this study, the bottom-up hybrid approach is used to investigate the self-assembly of FFF along with the conformation, packing, and organization of the molecules within the aggregates. To that end, two CG mapping schemes are employed, which yield two CG models. The first scheme uses one bead to represent the aromatic ring, whereas the second model uses three beads. The rationale underlying the 3-bead model is to resolve the planarity and '2-Dimensional' excluded areas characteristic of aromatic rings. Each model yields a distinct self-assembled nanostructure, which agrees with experimental results. This difference is observed despite the fact that the models share both their underlying AA model and the CG model procedure by which salient AA features are extracted. This highlights the influence of molecular representation in bottom-up CG models. The 1-bead representation of the aromatic ring yields a solid core nanorod with a "layercake" arrangement. Within each layer, interactions between N-terminal and C-terminal beads suggest that the charged groups that they represent are important to early aggregation. The final aggregate demonstrates predominant interactions between the amide groups, possibly arising due to hydrogen bonding. In contrast, the 3-bead representation of

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

the aromatic ring yields a solid core nanosphere with the aromatic rings tightly packed inside the sphere. In this model, the side chains drive the formation of micelles which shield the hydrophobic side chains from the aqueous solvent. Over time, those micelles coalesce into larger spherical aggregates with amide interactions slowly forming at the surface.

METHODS

Mapping Schemes. Two CG schemes for the zwitterionic triphenylalanine (PHE) peptide (FFF) sequence are used. The CG mapping schemes are summarized in Figure 1. In the first

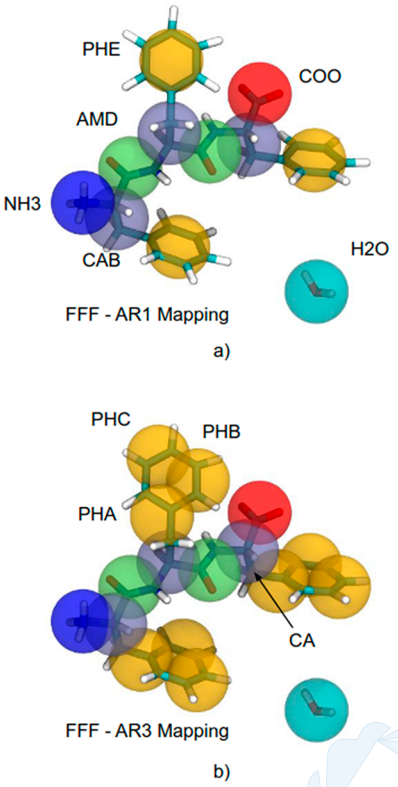


Figure 1. CG mapping schemes for the two models of triphenylalanine. Schemes depict AA molecules (solid lines), with transparent CG beads overlaid. Terminal and amide group beads are identical in the two schemes. (a) AR1 peptides comprise 10 beads of 5 types. Water is a single bead. (b) AR3 peptides comprise 16 beads of 7 types. Water is the same as in AR1.

CG scheme (AR1), five bead types are used: backbone carbon; amide group; N-terminal; C-terminal; and aromatic side chain. Each peptide molecule in the AR1 model comprises ten beads: three backbone, two amide groups, three side chains, one for C-terminal and one for N-terminal. In the second CG scheme (AR3), each side chain is represented by three beads. Hence, each peptide molecule in the AR3 model has 16 beads.

The mass of each CG bead is equivalent to the sum of the masses of its constituent atoms. Also, the center of mass of each CG bead is equivalent to the center of mass of its constituent atoms. Water is represented as a single bead (Figure 1). No CG bead carries an explicit charge. Since the peptide sequence in this study has a zero net charge, counterions are not required to maintain charge neutrality of the system.

Codes and simulation input files developed to run simulations, develop CG potentials, and analyze the results are available in a Github repository.[35] All molecular visualizations are prepared using VMD.[36]

All Atom Simulations. A set of reference AA MD simulations are performed at several peptide concentrations (see Supporting Information (SI) Table SI.1) in aqueous solution to enable sampling of equilibrium configurations for the development of the CG potentials and its validation. The AA MD simulations are run using the AMBER99 force field[37] which is chosen for its accurate resolution of peptide conformations[37] and solvation in water. Water is explicitly modeled using the three-point SPC/E model.[38] Prior to sampling MD trajectories (using production simulations), energy minimization, and equilibration simulations are performed. The production MD simulations sample the NPT ensemble. A cubic box with a 4 nm edge length and periodic boundaries is used except where noted. For the initial configuration in each system, the peptide and water molecules are randomly distributed in the simulation box. The temperature is set to 300 K via the Nose–Hoover thermostat. The system pressure is set to 1.0 bar via the Parrinello–Rahman barostat. The Lennard–Jones cutoff is set to 1.2 nm with long-range dispersion correction for energy and pressure applied. The particle mesh Ewald (PME) method[39] is used to calculate electrostatics with a cutoff of 1.2 nm. All bonds are restrained with the LINCS[40] algorithm. The MD calculations are performed using a leapfrog integrator and a time step of 2 fs via the MD simulation package GROMACS 2018.[41−44]

Coarse-Grained Potential Development. The CG beads interact with each other via CG bonded and nonbonded potentials. The bonded potentials capture the interactions associated with bonds, angles, and proper dihedrals. The nonbonded potentials resolve the interpeptide, peptide–water, and water–water interactions.

Boltzmann Inversion. Boltzmann inversion (BI) is a popular technique for deriving a potential energy function from a corresponding structural feature of a reference system. To provide such a reference, a pairwise radial distribution function (RDF) of a target degree of freedom (DOF) is derived by mapping a well-equilibrated AA trajectory into its CG representation and sampling the pairwise distances. The resulting reference distributions are inverted by BI using eqs 1−4[27,30,31] to obtain tabulated CG potentials U for bonds d, angles θ, dihedrals φ, and interparticle distances r, respectively. The potentials are state dependent and so depend on the temperature, here measured in Kelvin and scaled by the Boltzmann constant k_B.

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

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

$$U^{CG}(\phi, T) = -k_B T \ln(P^{CG}(\phi, T)) \quad (3)$$

$$U^{CG}(r, T) = -k_B T \ln(g(r, T)) \quad (4)$$

Iterative Boltzmann Inversion. Iterative Boltzmann Inversion (IBI) is a corrective scheme in which an input potential is applied to a simulation from which a sample of structural measures is calculated. The difference between the simulated structural measures and the reference is calculated, and the input potential is recalculated as shown in eq 5. It has

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

been shown that successive iterations of IBI are capable of efficiently correcting BI potentials[27] to better approximate a target structural measure, for example, an RDF.

$$V_{i+1}(\rho) = V_i(\rho) + k_B T \ln\left(\frac{P_i(\rho, T)}{P^{ref}(\rho, T)}\right) \quad (5)$$

Force Matching. Force matching (FM) is another scheme used to extract CG potential functions from AA MD reference trajectories.[28] Rather than targeting a specific structure, FM projects the multibody AA forces onto the CG-mapped representation.[45] Differences in the CG force relative to the net AA force are minimized using least-squares, as shown in eq 6, resulting in a system of linear equations. With a suitable number of frames from an AA reference trajectory, these equations are overdetermined and can be solved to approximate the potential of mean force (PMF). Typically, several blocks of frames of trajectories are sampled, with their PMFs solved and averaged to produce the final potential function.

$$\chi^2 = \sum_m^M \sum_n^N |F_{mn}^{ref} - F_{mn}^{CG}|^2 \quad (6)$$

In this study, an extension to the FM technique in which all bonded and nonbonded intrapeptide interactions are excluded from the calculation (that is, FM with exclusions)[46] is used. The FM with exclusions method samples only the interpeptide interactions. Models using this technique have been capable of recovering the nonbonded pairwise RDFs of reference AA systems.[20]

Sampling Reference AA Trajectory. The Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA)[47] version 1.6 contains implementations of BI, IBI, and FM. The VOTCA package is used for the development of the CG potentials, mapping to CG representation, and RDF calculations.

For BI, coordinates from a sample of 250,000 frames of AA MD simulation of a single peptide in water is used to produce smooth target RDFs. IBI is applied successively to correct the resulting potentials.

The FM method uses forces calculated from a sample of 50,000 frames of AA MD simulation for 25 peptides solvated in aqueous solution in a cubic simulation box of length 4 nm. In practice, FM calculations are limited by the number of AA reference frames that can be held in memory for computation of the forces. Therefore, several blocks of FM are performed with 1000 frames per block. The VOTCA implementation of FM uses cubic splines to fit force data across all blocks. The number of splines used is a function of the distance over which FM is performed and the grid spacing in that range. The range and grid spacing are set such that an average of 14 splines per interaction are used.

Coarse-Grained Simulations. A frame from the equilibrated AA reference trajectory is mapped to its corresponding CG representation to provide an initial configuration for the CG simulation. The CG simulations are run at constant volume, with a box length of approximately 4 nm set to match those from the "equilibrated" AA system. Tabulated potentials presented in the Results and Discussion section are used for the CG simulations. The potential energy cutoff is set to 1.2 nm. This is consistent with the minimum image convention in which pairwise interactions are truncated to prevent a particle from interacting with its own periodic image.[48] The MD calculations are performed using a leapfrog stochastic dynamics integrator with a time step of 1 fs via the MD package GROMACS 2018.[41−44] Temperature is set to 300 K via the stochastic integrator with an inverse friction coefficient of 1 ps.

Backmapping and Backmapped-Atomistic Simulations. The method presented by Wassenaar et al.[49] has been adapted to project AA coordinates onto CG coordinates, creating a backmapped atomistic configuration. The backmapped configuration is energy minimized excluding nonbonded interactions and then energy minimized again with no exclusions. Four consecutive, short, position-restrained NVT simulations are run in which the time step is set to 0.2, 0.5, 1, and 2 fs, respectively. A final, short MD simulation results in a stable backmapped AA configuration. Following these steps, the system is simulated for 10 million iterations using a time step of 2 fs and the NPT ensemble. Root mean square deviation (RMSD) of all peptide atoms is measured with respect to the initial relaxed frame of the simulation.

The final CG configurations for both models for the 1, 8, and 32 peptide systems are backmapped. After the energy minimization routine, each backmapped configuration is simulated using the same parameters as those corresponding to the AA reference. The RMSD of all peptide atoms in the simulated trajectory is measured relative to the initial backmapped configuration. The AA reference trajectories are sampled from 300 to 320 ns of the production simulation in the respective systems.

RESULTS AND DISCUSSION

Water−Water Potentials. The CG water−water potentials are derived from AA trajectories using the AA water−water RDF. The AA trajectories are obtained from an AA MD simulation run for 200 ns using a time step of 2 fs and sampled at intervals of 2 ps (yielding 100,000 frames). The water−water RDF is calculated using the AA trajectory and inverted using BI to yield an initial estimate of the CG water−water potential.

A CG MD simulation of a pure water system is run for 200 ps using a time step of 2 fs and sampled at 200 fs intervals. Details of the CG MD simulations are provided in Methods. The first 20 samples of the CG MD trajectory are discarded, and the remaining samples are used to calculate the CG RDF. The CG water−water RDF is compared to the AA water−water RDF, and IBI is used to refine the CG water−water potential. After 299 iterations of IBI, the CG RDF matches the reference AA RDF, thereby yielding the final CG water−water potential.

The potential derived in this manner reproduces the water−water RDF in box-of-water simulations as well as when a single peptide solute is introduced (Figure S2 and Figure S9). The CG water−water potential is derived by only using interactions between water molecules and is therefore transferable over some suitably low range of peptide concentrations.[20] Furthermore, it is suitable for CG models that use the same coarse-graining scheme for water as the two models in this study. Conversely, this potential may not be suitable for higher solute concentrations or systems that have a significant impact on the local packing of water.

Initial Bonded Potentials. The initial potentials for bonds and angles are generated via BI. An AA MD simulation of a single peptide solvated in water is run for 500 ns using a time step of 2 fs and is sampled at a rate of 2 ps (yielding 250,000

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

frames). The frames are used to calculate the CG-mapped RDFs corresponding to the bonded interactions for each CG model. The BI method is used to invert these RDFs and provide an initial set of potentials to maintain the rudimentary intrapeptide structure of a single solvated peptide in the CG MD simulation.

These initial BI potentials, derived from a single peptide in water, exclude information about interpeptide interactions. Since one of the goals of the study is to examine the self-assembly of peptides in aqueous solution, the simulations of interest will have multiple peptides and water molecules and require intermolecular nonbonded potentials. These nonbonded potentials will affect the conformation of a single peptide in solution, causing a change in the bonded parameter distributions. Therefore, for aggregation studies, the BI potentials must be refined in the presence of interpeptide interactions. This refinement is achieved using the iterative process described in Refined Bonded Potentials below.

Peptide−Water Potentials. Peptide−water potentials are derived by using IBI. The AA peptide−water RDFs are calculated using a trajectory sampled from an AA simulation of a single peptide solvated in an aqueous medium. The simulation spans a duration of 500 ns and is sampled at a rate of 2 ps (yielding 250 K frames). The AA trajectory is inverted by BI to yield the initial potentials. IBI iterations are based on CG simulations spanning 5 ns and sampled at 1 ps intervals (5,000 frames). No dihedral potentials are applied in the CG simulations.

An AA solvation RDF derived from a single peptide in water is set as the IBI target. The IBI potentials implicitly depend on the sampled peptide structure, which leads to a somewhat slow convergence of iterated CG RDFs. The peptide−water RDFs obtained after successive iterations of IBI are compared with the AA reference. This is to aid the selection of a potential that maximizes their qualitative similarity without excessive refinements and therefore overfitting. The potentials were selected after 99 iterations of IBI for model AR1 and after 69 iterations for model AR3.

Refined Bonded Potentials. A CG simulation using all potentials that have been developed at this point is run for 5 ns and sampled at an interval of 2 ps (yielding 5000 frames). Distributions are calculated for the bonds and angles. Bond distributions agree well with corresponding measurements from the AA reference, so the potentials for the bonds are excluded from further refinement. For both the angles and dihedrals, refined IBI potentials are chosen which produce CG structural distributions whose peak locations agree with those seen in AA reference (see Figure S1 and Figure S8). Applying IBI to only a subset of potentials means that one must accept a trade-off between improving the target distributions and negatively impacting some of the other structural measures. In practice, the selected number of iterations yields the maximum possible improvement in the behavior of the structural target without noticeably degrading other parametrized degrees of freedom.

Existing angle potentials are refined using IBI. The iterations are based on a CG simulation spanning 20 ns which is sampled at an interval of 2 ps (yielding 10,000 frames). Convergence of the angle distributions is much quicker than in the case of the nonbonded peptide−water potentials, as the angle potentials depend only on the internal structure of the peptide. The angle potentials are refined by three iterations of IBI.

Potentials for dihedrals are introduced at this step. Initial estimates of the distributions are calculated using the AA simulation of a single peptide solvated in water spanning 500 ns (as previously described in Initial Bonded Potentials). Potentials are selected after one iteration of IBI, using the same CG simulation parameters as those described for angles.

Peptide−Peptide Potentials. Nonbonded peptide−peptide potentials are derived using the FM with exclusions method as described in Methods. These potentials are derived from AA trajectories that are sampled from an AA MD simulation encompassing 25 peptides and 1573 water molecules (see Table SI.1). The AA MD simulation uses a time step of 2 fs and is run for 500 ns. The final 200 ns of the MD simulation is sampled at the rate of 2 ps to yield 100,000 frames which are used to derive the potentials via the FM with exclusions method.

The FM with exclusions method removes the impact of nonbonded intrapeptide interactions, which are represented in the AA reference system. It has been noted that the resulting CG potentials may therefore fail to replicate the intramolecular conformations of their respective AA counterparts.[18] It should also be noted that FM potentials will depend on the sample of many-body interactions in the target AA trajectory, implicitly limiting the transferability of such potentials.

Time Scales and Computational Efficiency. The utility of a CG model depends on its ability to accelerate system dynamics. Further, the model should scale efficiently with computational resources. Therefore, the performance of the two CG models is examined by using two measures.

The first measure is the diffusion coefficient of a single isolated peptide. The diffusion coefficient is an effective characteristic of the dynamics of a system and, therefore, may be used to compare different systems. The mean square displacement (MSD) of the peptide is measured from the MD trajectories, and the Einstein relation is used to calculate the diffusion coefficients. Using this approach, the diffusion coefficients of a single, isolated peptide in the AA, AR1, and AR3 models are, respectively, 0.2431, 2.2079, and 0.9175 × 10^{-5} cm^2/s. The speedup in dynamics of a CG model is given by the ratio of the diffusion coefficients of the single isolated peptide in the CG to AA representation. This ratio is measured to be 9.08 and 3.77, respectively, for AR1 and AR3 models. This is a somewhat coarse measure of efficiency, since single-peptide diffusion coefficient measurements exclude the role of interpeptide interactions, which are an important contributor to the dynamics of multipeptide systems.

The ultimate goal in developing a CG model is to use it for efficiently examining the behavior of large systems (namely, multiple peptides) over extended spatiotemporal scales. This acceleration is achieved by producing a smoother potential energy landscape than what is observed for AA models.[50,51] This difference in smoothness implies that the AA and CG time scales are not directly comparable. Instead, the acceleration of the particle motion by the CG system and execution of enough CG MD cycles to achieve an equilibrium state far beyond what is accessible to AA is validated. This will require the efficient use of high-performance computing resources. Hence, the second measure of computational efficiency is the core performance scaling of each CG representation relative to the AA representation.

For the second measure, the model is tested for strong scaling in which doubling the number of processors halves the compute time required to solve a problem of a fixed size. To

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

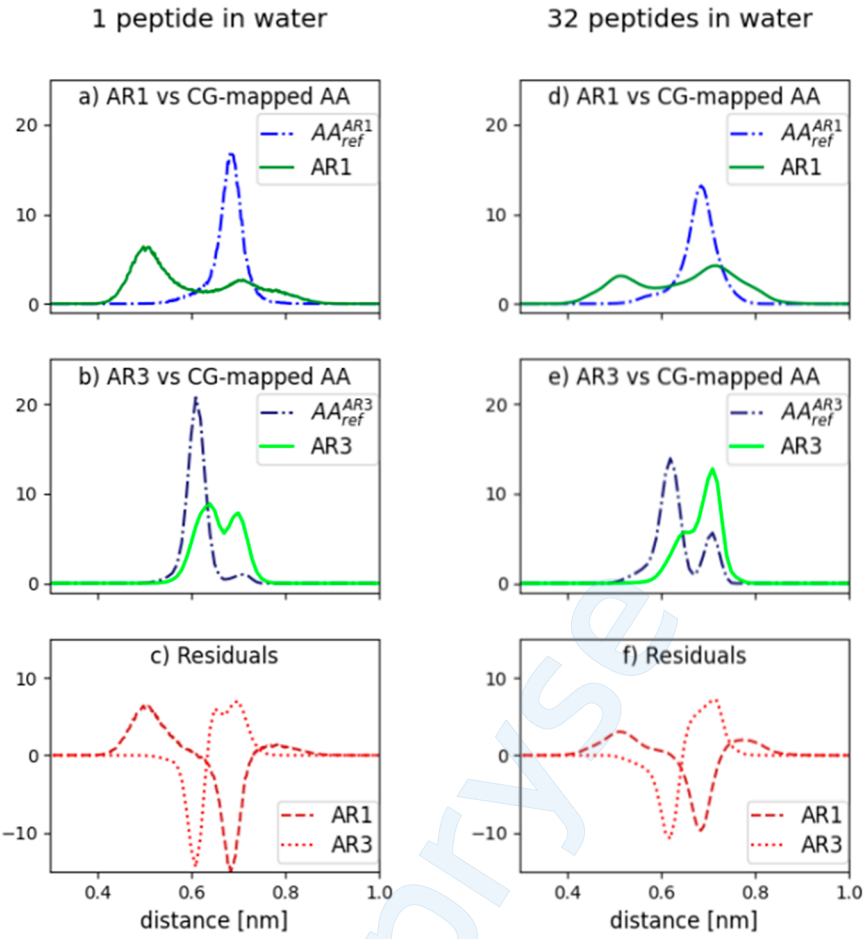


Figure 2. Distributions of peptide backbone end-to-end distances calculated from simulations of the AR1 and AR3 models. The left column reflects systems of a single peptide, and the right column reflects systems of 32 peptides. In the upper panels (a, b, d, e) the CG model results are compared with AA reference data. AA reference coordinates are mapped into the appropriate CG representation. It can be seen that the AA reference data produce different distributions for each of the two mapping schemes, owing to the difference in the definition of the backbone beads (see Figure 1 and SI file). The bottom panels (c, f) show residual values, calculated as the difference between the simulated CG results and the respective CG-mapped AA results.

test this, 8 peptides are solvated in a simulation box (of dimension 4 nm) with 2000 water molecules in the AA, AR1, and AR3 model representations. Each system is simulated for 2 million iterations (that is, at time steps of 2 and 1 fs, respectively for the AA and two CG representations). The core performance is measured in nanoseconds per day, as reported by the simulation package. Test simulations are performed using 1, 2, 4, and 8 cores on a Regular Memory node of the Pittsburgh Supercomputing Center Bridges-2 cluster.[52,53] The results from these tests are summarized in Figure S16, showing that both CG models exhibit strong scaling up to 8 CPU cores.

Structure of a Single CG Peptide. The fully parametrized AR1 and AR3 models are each used to simulate a single peptide solvated in water, to examine the behavior of parametrized degrees of freedom and overall peptide conformation. Distributions of the DOFs which are explicitly parametrized in each model generally sample the same ranges as the AA reference, with most DOFs also correctly resolving the locations of peaks in the reference distribution. It is noted that the smooth CG potentials may introduce biases in the sampled configurational ensemble, leading to peak heights of structural distributions that differ from the AA reference. While such deviations are not necessarily unphysical,[30] they do suggest a CG model which imperfectly captures the details of the underlying AA system. Some instances are observed in which parametrized CG angles are able to take on values not represented in the AA reference. See Figures S1 and S8 and Overview of Calculated Structural Distributions in the SI. These relatively small deviations emerge from the over-constraint of the internal peptide geometry and therefore may be seen as a consequence of the parametrization scheme.

To assess whether the models give rise to a plausible overall conformation, peptide length and mass distribution are calculated. The length of the peptides is characterized by the end-to-end distance, which is given by the distance from the center of mass (CoM) of the first to the CoM of the third backbone bead. The mass distribution is characterized by the radius of gyration (R_g), as determined by the weighted spatial organization of the particles. The AA reference data are CG-mapped for comparison of observables from the two models.

End-to-end distance distributions for a single peptide in each model are shown in Figure 2a–c. In the AR1 model, the backbone beads include both the α and β carbon, whereas in the second model, only the α carbon is used. The AR1 mapping of the AA reference data shows a single peak in the distribution of end-to-end distances, whereas the AR3 mapping shows two peaks. The AR1 model produces a substantial proportion of peptides that appear shorter than those seen in

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

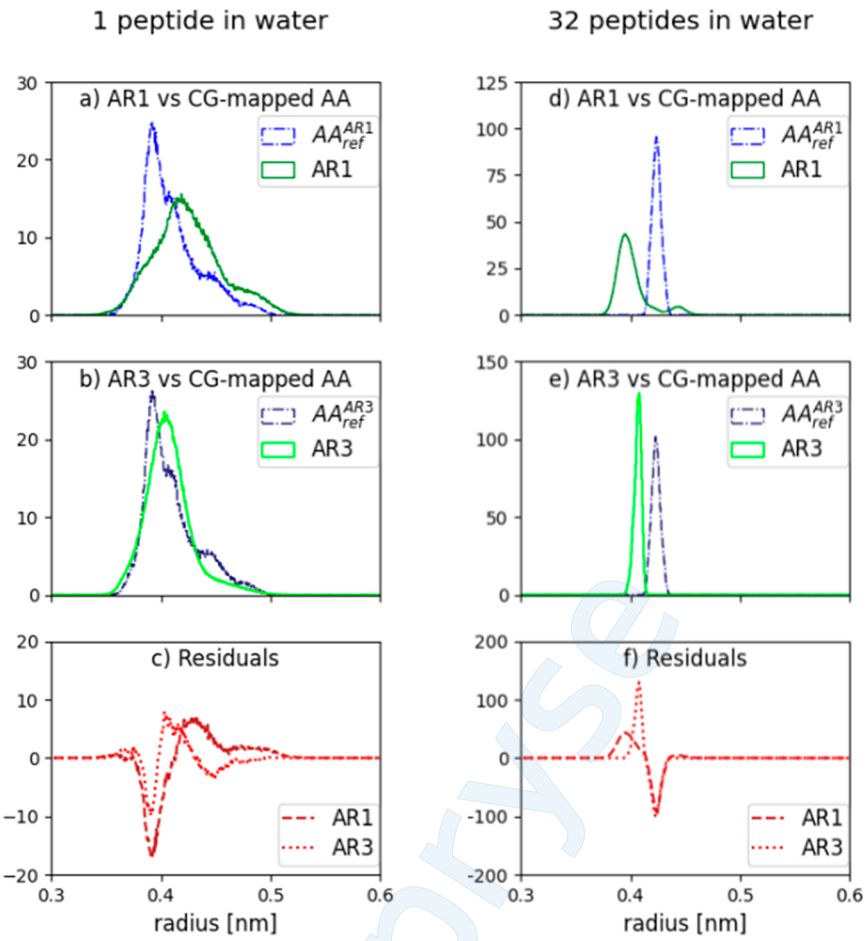


Figure 3. Distributions of peptide radius of gyration (R_g), calculated from simulations of the AR1 and AR3 models. The left column reflects systems of a single peptide, the right column reflects systems of 32 peptides. In the upper panels (a, b, d, e) the CG model results are compared with AA reference data. AA reference coordinates are mapped into the appropriate CG representation. R_g calculations based on the AA reference data produce nearly identical distributions when they are mapped into each of the two mapping schemes. The bottom panels (c, f) show residual values, calculated as the difference between the simulated CG results and the respective CG-mapped AA results.

the AA reference data and a somewhat smaller sample of peptides that are longer. These deviations are apparently due to the model's biased sampling of the angle distributions (see Figure S1). The AR3 model performs better, with the end-to-end distance sampling the same range as the AA reference data and recovering the same peaks of the distribution. Peptides from both models closely reproduce the range of R_g values observed in AA reference data (see Figure 3a–c).

The appearance of unexpectedly short or long conformations in the AR1 model suggests that its parametrization is insufficient to produce a rigid conformation. Despite a similar parametrization strategy, the AR3 model produces peptide lengths comparable to the AA reference data. This suggests a role for steric interference of the AR3 side chain beads in extended conformations, a factor that cannot be replicated by the single-bead side chains of the AR1 model.

Systems with Multiple Peptides. Both fully parametrized models are also used to simulate systems of 2, 4, 8, 16, and 32 peptides solvated in water. Individual peptide structures are checked to ensure consistency with corresponding characteristics resulting from the AA reference systems. The formation of multipeptide solid nanostructures is observed in the 32 peptide simulations for both models. The formation of nanostructures via self-assembly is inaccessible via AA MD simulations, primarily due to extended relaxation times corresponding to the association of peptides into tightly packed formations, which are characteristic of equilibrium nanostructures. The CG MD simulations are able to resolve the self-assembly of these structures very quickly, owing to the smoother energy landscapes corresponding to the CG force field. Hence, the self-assembled nanostructures from the CG simulations are compared to the corresponding results from experiments to validate the predictions. Interparticle contacts (or interactions) in CG and backmapped atomistic-scale trajectories are used to investigate the contribution of the CG groups of atoms to the final nanostructures.

Peptide lengths in multipeptide simulations with the AR1 model show a pattern of deviation similar to that observed in the single peptide case (see Figure 2 and Figure S6). This behavior is mitigated in assemblies, where the mean of the CG length distribution is shifted toward that of the AA reference.

For the AR3 model, the distribution of lengths has better qualitative agreement with the corresponding results in the AA reference, reproducing the peak locations and the correct range of the distribution at all concentrations. The AR3 model tends to yield end-to-end distances at the high end of the distances sampled by the AA reference. This contrasts with the low end of the distances sampled by the AR1 model. Comparison of the distribution of the end-to-end distances from the CG models

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

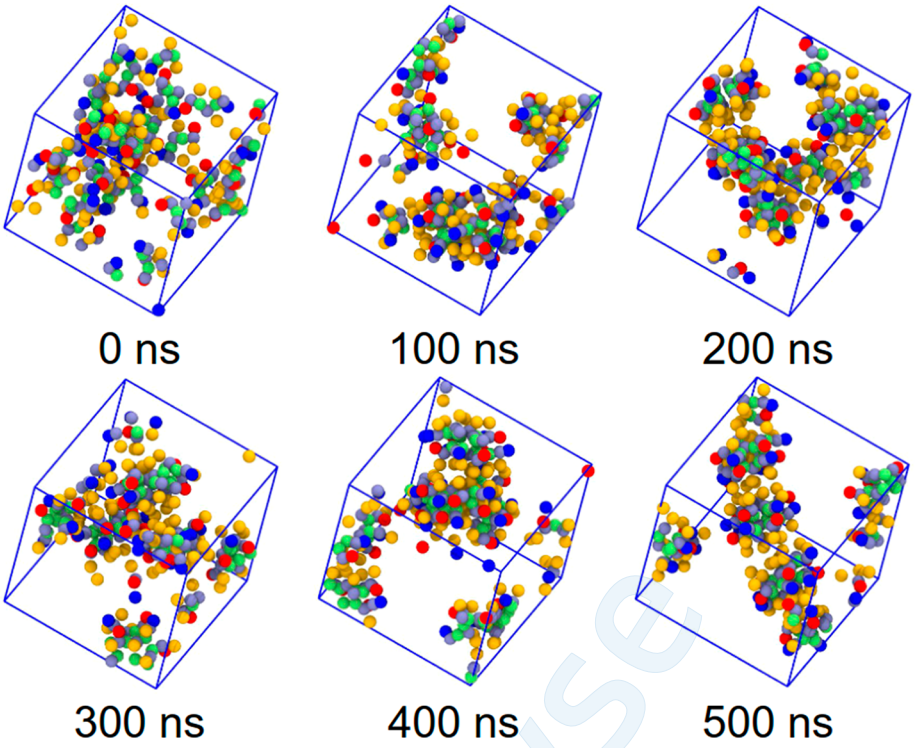


Figure 4. Process of self-assembly of 32 peptides in simulation with the AR1 model yields a solid nanorod.

with those corresponding to the AA reference is shown in Figure 2.

Both models yield R_g values consistent with corresponding values from the AA reference when simulating a single peptide in solution (Figure 3). At a high concentration of peptides, both models yield values of R_g that are significantly lower in the CG representation than in the AA representation, despite the differences in overall peptide length observed for the two models. This result arises due to rearrangement of the peptide into a nearly planar conformation in which side chains project from the backbone in the same direction to minimize their interactions with water. The associations between the side chains within the aggregate help overcome the energetic barrier to this cis-conformation of the side chains, which in turn reduces the R_g.

The distributions of the bonded degrees of freedom resulting from the trajectories for both the AR1 and AR3 models closely match the corresponding AA reference (Figure S1 and Figure S8). For the AR1 model, the end-to-end length systematically deviates from the AA reference, allowing peptides to adopt very short or very long conformations (Figure 2). The end-to-end length of the peptide is not a parametrized property in either model and is treated as an emergent property which is a function of the bonded potentials. Viewing the deviation of the end-to-end distance for the AR1 model from the corresponding result for the AA reference as an error, the AA model is surmised to exhibit coupling of the bonded degrees of freedom which is not captured by the CG model. Given that this error is mitigated in multipeptide assemblies (see Figure S3 and Figure S6)—and that it is ameliorated in the AR3 model—it is hypothesized that the extended peptide conformation is mediated by the steric effects of the bulky side chains. The exact nature of these effects within the CG models bears further study, but the key assumption is that nonbonded interactions of the side chains, with each other and with solvent, drive the stability of the extended peptide conformation.

Interpeptide RDFs show that the CG model samples the same ranges as the AA reference, with broad agreement overall in the location of distribution peaks (Figure S5 and Figure S11). There are however notable deviations reflecting the degree of aggregation at higher concentrations. In the AR1 CG system of 32 peptides, the terminal beads do not come into close proximity with beads of the same type (COO−COO and NH3−NH3) at the same frequencies seen in AA reference data. This is explained by the regular antiparallel arrangement of peptides seen in the final aggregate structure described below. In contrast, close interactions of the amide groups (AMD−AMD) are very highly sampled in the CG representation despite being largely absent from the AA representation. One possible explanation is that amide groups in the AA model are strongly attracted, but that the many-body nature of the AA representation results in a high free energy barrier to their close approach, a barrier which is incorrectly eliminated by the CG model.

Similarly, in the AR3 CG system of 32 peptides, interpeptide RDFs sample ranges and peak locations similar to those of the AA reference data, often at similar frequencies. Notably, terminal beads (NH3 and COO) tend not to approach other beads as closely in CG representation as they do in AA representation. This is once again explained by the retreat of the terminal beads to the surface of the final aggregates. The amide groups in the AR3 model also show closer proximity compared to the AA reference. Once again, this may be explained by CG forces which represent a subset of the physics associated with the AA reference but which are limited in reproducing multibody effects.

Self-Assembly of Triphenylalanine into Nanostructures. The jagged potential energy surface of the AA reference system makes it impossible to resolve the assembly of multiple

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

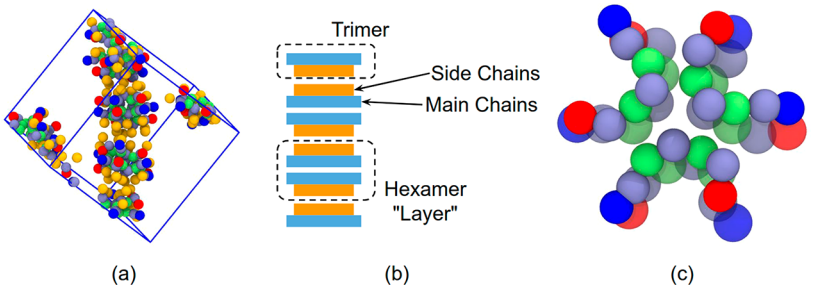


Figure 5. Nanorod observed in CG simulation with the AR1 model. (a) Side view of nanorod with all peptide beads in system. Bead colors as follows: Yellow - side chain; Blue - N-terminal; Red - C-terminal; Purple - backbone; Green - amide. (b) Schematic of nanorod stacking. Peptide trimers formed with main chain beads in planar arrangement, and side chain beads in a parallel plane. (c) Main chain and terminal beads of two adjacent peptide trimers, viewed looking down normal to trimer plane (i.e., with nanorod longitude). Side chain beads are omitted for clarity. The peptide beads in the lower trimer plane shown larger and transparent.

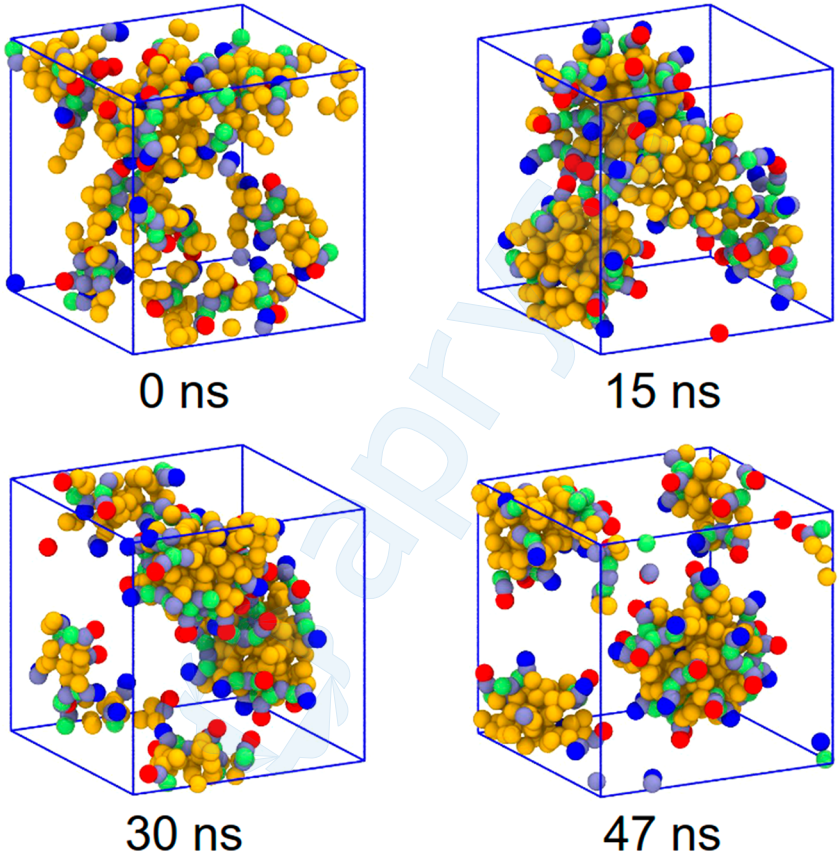


Figure 6. Process of self-assembly of 32 peptides in simulation with the AR3 model yields two solid nanospheres each roughly 3 nm in diameter.

peptides, which is a primary motivation for the development of a CG force field. The bottom-up CG model smoothly approximates the AA reference force field, subject to the configurational ensemble sampled during CG force field development. One commonly used method to validate a bottom-up CG force field is to compare the derived CG potentials to the potential of mean force (PMF) calculated from the AA reference.[30,31,33,54] However, the CG model is generally used to explore very large spatiotemporal scales that are far from the conditions under which the AA potential energy surface is initially developed. Therefore, although qualitative agreement of the PMF in small systems is a reasonable desideratum, it is not necessarily a very informative measure of the quality of a model for large systems. It is for this reason that rather than evaluating the PMF, a direct comparison is made with experimentally observed nanostructures, which is considered as the second way to calibrate the model. Hence, one relies upon the rigorous design of the FM procedure to produce a CG potential energy surface that at least qualitatively mimics the underlying AA potential energy surface.

The AR1 model was used to resolve the assembly of 32 peptides initially dispersed in an aqueous solvent. Ten independent simulations were run ranging from 40 million to 500 million iterations. Aggregation is characterized by rapid association of small numbers of peptides into trimers and hexamers and then slow association of these oligomers into a larger aggregate. In three simulations, each spanning 100 million iterations or more, a stable solid core nanorod is formed. Figure 4 shows the process of forming a nanorod by

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

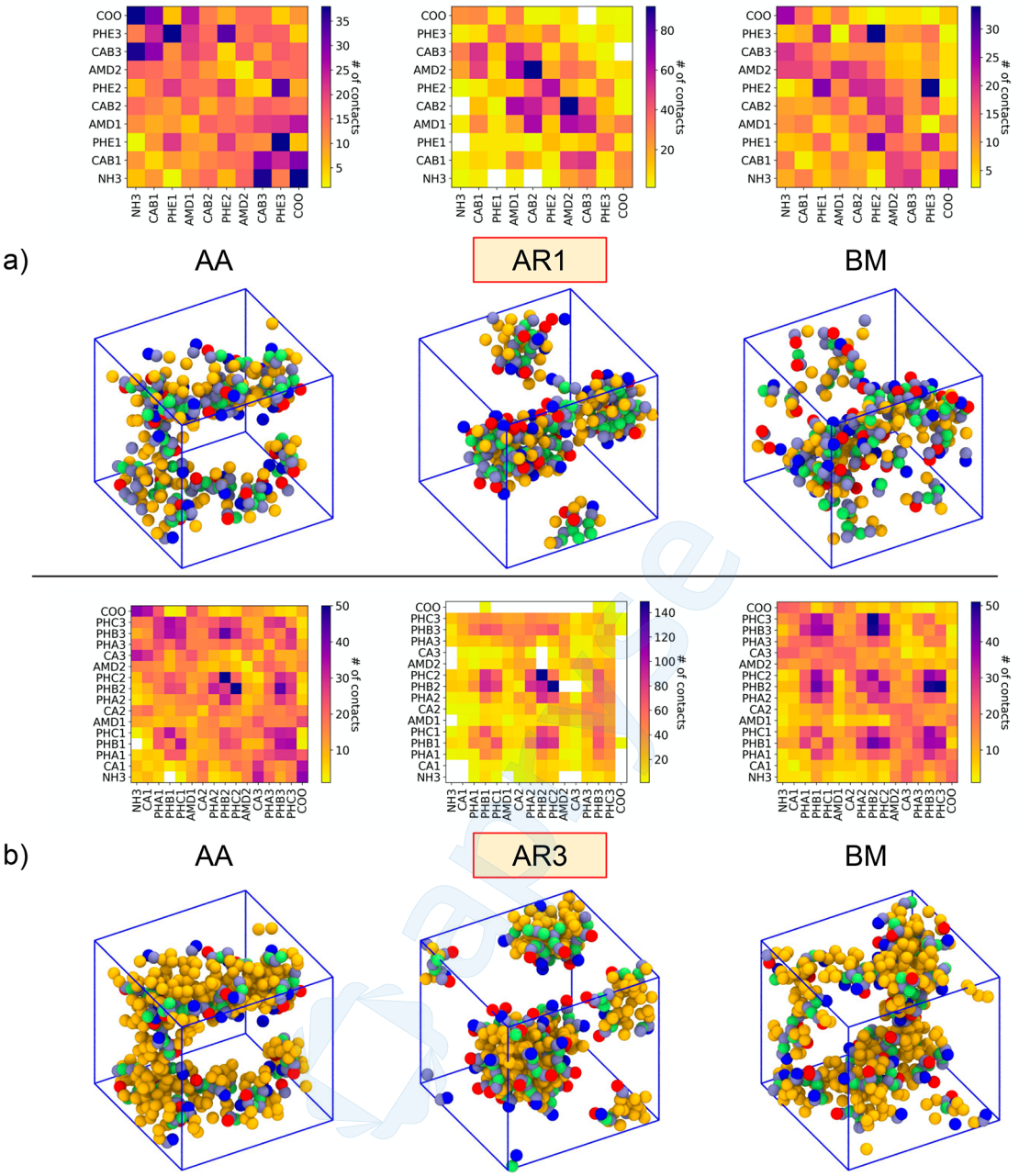


Figure 7. Contact maps of 32 peptides in AA, CG, and backmapped models, with visualization of each system. (a) Simulation frames in the AR1 mapping scheme. (b) Simulation frames in the AR3 mapping scheme. A contact is defined as a pairwise distance of less than 0.6 nm.

simulation spanning the longest duration. Individual peptide lengths never exceed 1 nm (Figure S3). Therefore, in a 4 nm simulation box with a 1.2 nm potential energy cutoff, no single peptide will interact with its periodic image. This is a stronger condition than the minimum image convention and is worth noting, especially in this case where the resolved nanostructure encounters the periodic boundary.

The nanorod encompasses multiple layers of peptides with each layer encompassing 5 to 7 peptides. In each layer, the side chain beads of the peptides are shielded from water, while the main chain beads form the interface between consecutive layers. The interfaces between the layers in the nanorods include amide groups that are in close proximity to one another. A view of a typical half-layer trimer is shown in Figure 5, where three peptides are distributed with radial symmetry in the plane of the layer, in close contact with an adjacent trimer. This result is consistent with elongated structures reported by an earlier experimental study.[10] That study reported that the elongated structures contained a high β sheet content, which the authors suggest is a result of close associations of amide groups in the observed nanostructures.

Using the AR3 model, ten independent simulations of 32 peptides were run ranging from 20 million to 47 million iterations. Aggregation in this model typically occurs with an initial rapid association of all peptides into a single aggregate. This aggregate breaks into smaller nanospheres composed of roughly 10 or more individual peptides. In a simulation spanning the longest duration, the self-assembly of 32 peptides in aqueous solution yielded two spheres of approximately the same dimensions within 40 million iterations (Figure 6). An

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

examination of the self-assembly pathway demonstrates the initial formation of irregular elongated structures, which evolved into two spheres around 25 million iterations. The self-assembled nanospheres are also consistent with an earlier experimental study of the self-assembly of Boc-FFF peptides into nanostructures.[13] The authors of that study posited side chain packing as a likely driver of the dense spherical structures, but they were unable to discern a mechanism.

Backmapping. The role of backmapping in this study is twofold. First, it allows verification that the AA details can be reinstated without any numerical instability. Both models in this study passed this test. Second, the reintroduction of AA degrees of freedom provides a method by which one can evaluate the extent to which the CG model faithfully approximates the AA reference at spatiotemporal scales where AA MD simulations are cost-prohibitive. For this, one checks for signs of stability in AA simulations of the backmapped systems.

The ordered nanostructures obtained from both CG models quickly dissolve when the systems are backmapped and allowed to evolve in the AA representations. One explanation for this instability is that the many-body AA representation contains details which locally destabilize the nanostructures and that the CG model fails to capture these details. This is a strong possibility, since the production of the CG model necessarily averages out many local effects, especially within individual CG beads. Similarly, the FM approach used in this study derives two-body nonbonded parameters that only implicitly account for the many-body structures seen in the AA reference.

Having failed the check for stability in the backmapped models, it is useful to further explore which features of the CG model seem to be in agreement with the AA reference and which do not. To do this, AA contacts that persist when the backmapped system is allowed to evolve under the Amber force field were further investigated, as discussed below.

Contact Maps and Cluster Counts. Interpeptide contact maps for the 32 peptide systems are used to estimate the predominant interactions between the molecules in an aggregate. To generate these maps, the last frame from the AA reference, CG, and backmapped (BM) simulations are extracted. The AA and BM configurations are mapped to the CG representation, and contact maps are calculated.

Figure 7a shows representative contact maps for the AR1 model. The AA contact map shows dark areas at the antidiagonal corners of the map, reflecting interactions between the charged end groups. There are also dark patches for side chain interactions between the first and third side chain groups and second and third side chain groups. This suggests an end-on-end arrangement of consecutive peptides in which the first two side chain groups of the first peptide interact with the third side chain group of the second peptide.

The CG contact map shows much greater interactions adjacent to the second PHE residue as well as a continuing preference for antiparallel arrangement. The most prevalent contacts are between antiparallel amide groups and their adjacent backbone carbons. This seems to imply that the backbone interactions stabilize the nanorod configuration.

The BM contact map shows a clear preference for interactions involving the second side chain. Interactions between the first and third side chain groups remain less common in the BM representation than in the AA representation.

Figure 7b shows contact maps for the AR3 model. The AA reference shows dark patches for the main diagonal and antidiagonal side chain interactions. With the AR3 mapping scheme, however, the finer differences in the nature of side chain interactions can be discerned. For example, interactions between first side chain groups are clearly dominated by the exterior beads, suggesting a configuration in which the outer tips of the rings are in proximity. Conversely, interactions between the first and third side chain groups are much more uniform over the three beads, suggesting a more planar configuration of the side chains.

The contact map of the CG nanosphere shows many close contacts and sharper peaks for the side chain interactions. The interactions for the first and third side chain groups are reduced. Moreover, the interactions between the third side chain groups are drastically reduced as the C-terminal bead retreats away from the spherical aggregate, with an analogous decrease in interactions for the first side chain groups due to the retreat of the N-terminal.

The BM system shows an even more pronounced preference for side chain interactions. This is attributed to the smooth CG energy landscape, which enables side chains to come into close proximity in a reasonable duration of time. This is not possible with AA MD simulations due to the prohibitively long relaxation times.

The AR3 model allows for a direct comparison of the angle between adjacent rings for the AA and CG representations of the system. The distribution of the angles for the AA representation shows peaks around 45° and 135°, corresponding to a "herringbone" arrangement. The CG configuration favors angles of 90° between aromatic rings, or "T-shaped" stacking. The angle distributions for the BM trajectories are qualitatively similar to those corresponding to the AA trajectories (Figure S14).

It is noted that neither the CG interbead contact propensity nor side chain orientation are specifically parametrized by the CG models. Thus, the packing and orientation of peptides present a valuable set of emergent characteristics of the CG models. The jagged AA potential energy landscape yields very slow dynamics once many peptides are in close proximity, which creates a high energetic barrier to the evolution of a tightly packed configuration, such as the ones seen in the CG models. Therefore, the aggregates found in the CG model cannot be validated by the AA model. In the future, we plan to run AA simulations with some enhanced sampling to enable validation for the CG model.

The contact maps and ring contact plots represent nonbonded interactions in the AA, CG, and BM systems. The higher maximum contact counts of the CG simulation contact maps correspond to the aggregated states and reflect distinct contact propensities that differ from those in the AA simulation. The BM simulations use the evolved CG simulated configurations as their initial condition. These systems are evolved for sufficient duration to demonstrate the numerical stability of these CG configurations under the AA force field. This confirms the plausibility of the simulated CG configurations.

The contact maps from BM simulations show certain similarities to their CG simulation counterparts. Comparison clearly shows that there are contact motifs resolved by the CG models that persist after short evolution of the dynamics of the corresponding AA system. It is surmised that similarities between the CG and BM contact maps indicate those packing

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

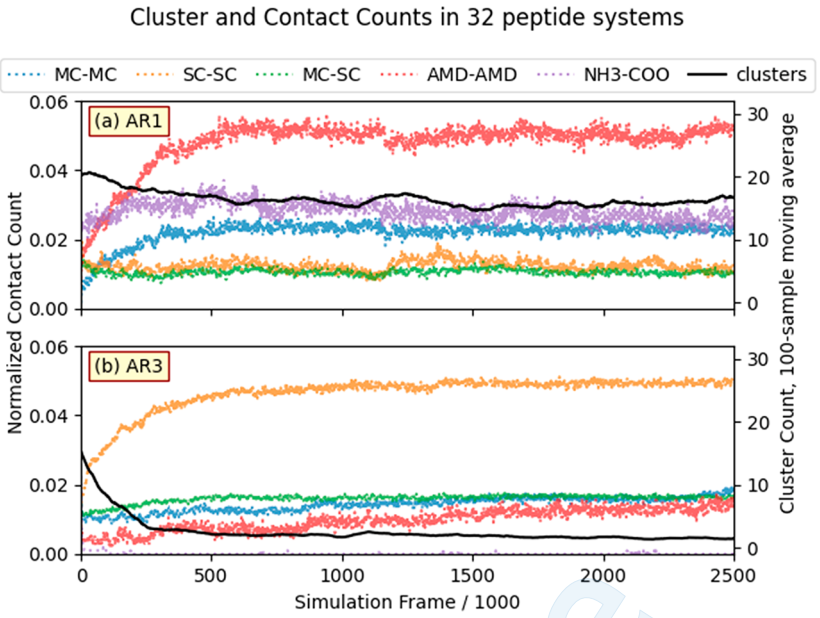


Figure 8. Time series plots of selected interaction contacts and cluster count. Left vertical axis indicates normalized contact count, which is the ratio of actual contacts to the maximum possible number of contacts for each select interaction. Right vertical axis indicates cluster count as a moving average. Horizontal axis truncated at 2.5 M MD steps to show trend. Clusters are defined as collections of peptides with pairwise distance of less than 0.35 nm between any constituent beads. (a) AR1 model results. AMD-AMD interactions rise sharply at the beginning of simulation, indicating an influential role in the nanorod aggregate. The relatively large cluster count in the AR1 nanorod is due to the relatively large spacing between consecutive layers of the aggregate, which nonetheless does not allow water to intervene near the axis of the nanorod. (b) AR3 model results. SC-SC interactions far outnumber all other interactions during nanosphere formation.

features of the CG model that are most consistent with the AA model. That is, those simulated CG contacts that do not rapidly dissociate in the BM simulation are the ones that can most stably be maintained by the AA model.

Dynamics of Self-Assembly. The dynamics of the assembly along with the predominant interactions driving the formation of the aggregates and their stability can be construed from measurements of the time evolution of the cluster count and the number of interactions between the different moieties. Figure 8 summarizes interaction counts for several bead types: backbone main chain (MC−MC); side chain (SC−SC); main chain to side chain (MC−SC); amide (AMD−AMD); and N-terminal to C-terminal (NH3−COO). Calculations of the cluster count were overlaid in each plot. These measurements were performed on 32 peptide systems using both the AR1 and AR3 models for a simulation spanning 20 million iterations.

In the AR1 model, the clustering process is initiated by interactions between the side chain groups closely followed by interactions between the terminal groups, leading to the formation of small platelet-like aggregates (Figure 4). However, as the molecules come into proximity (leveling out the interactions between the side chains and the terminal groups), there is a steady increase in the interactions between the main chain groups. When the cluster count reaches a steady value, the AR1 model shows the interactions between the amide groups followed by interactions between the terminal groups to dominate. This is followed by interactions between the main chains which are trailed by the interactions between the side chains. Hence, the amide, terminal, and main chain group associations play an important role in the organization of the approximately 15 aggregates present at the end of the simulation as the platelet-like aggregates further assemble to form a nanorod (Figure 8). The interface between the platelet-like aggregates within the nanorod is dominated by associations between the amide, main chain, and terminal groups. Furthermore, this result indicates a preference for antiparallel or end-to-end orientation of individual peptides within the aggregates.

In the AR3 model, the aggregation of the peptides is driven by interactions between the side chains with the formation of small micelles, which shield the side chain groups from the solvent (Figure 6). As the molecules come within proximity, the interactions between the main chains and those between the main and side chain groups begin to increase. The interactions between the amide groups grow at a slow and steady pace as the small micelles diffuse and coalesce to form larger sized micelles. After the cluster count reaches an approximate value of ∼2, the interactions between the side chains dominate and reach a steady value (Figure 8). This result is consistent with the tight packing between the side chains in the core of the nanosphere. The interactions between the amide groups and the main chain groups are less frequent than those observed for the AR1 model. In addition, the interactions between the terminal groups play virtually no role in the spherical aggregate.

CONCLUSIONS

Aromatic tripeptides have been observed to form regular aggregate structures in experiments. Previous simulation studies have focused on the formation of large-scale aggregates using top-down modeling methodologies or have described bottom-up methodologies which stop short of describing large nanoscale aggregates. In this study, two CG models are developed by using a bottom-up approach to elucidate conformation and interpeptide behavior of individual FFF peptides within assemblies.

The two models differ in their representation of the aromatic side chain as either one or three pseudo atoms. The extended

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

3-bead representation is chosen as a means to resolve the planarity and '2-Dimensional' excluded areas which are characteristic of aromatic ring structure. With the hybrid force field parametrization strategy described here, parameters of the model force fields are calculated from the trajectories of AA reference simulations.

Both of the resultant CG models resolve the solvation and structure of individual peptides in an aqueous solution. Each model also resolves the assembly of randomly dispersed peptides in an aqueous solution. These self-assembled nanostructures are consistent with the experimental observations. The 1-bead model produces solid core nanorod aggregates which are longitudinally extended in a "layercake" fashion. These nanorods show a high degree of association between the amide groups and the carbon backbones of neighboring peptides. The 3-bead model produces solid core nanospheres with tight packing of side chains in the interior of the sphere.

In both the nanorod and nanosphere assemblies, peptides conform to shield the hydrophobic side chains and expose the terminal groups to the solvent. The structure of the simulated nanorods suggests that the CG model has extracted information from AA reference trajectories, which leads to preferential interactions between amide groups, approximately analogous to hydrogen bonding. Likewise, the simulated nanosphere shows a highly efficient packing of aromatic rings not seen explicitly in the AA reference trajectories. These are two important examples of emergent characteristics captured by the CG models, which are derived using only AA structural information. It is further surmised that the 3-bead representation of aromatic side chains impacts the ability of the models to pack tightly, which bears further study.

The two models AR1 and AR3 adopt different representations but use the same reference data and methods in their force field development. This raises the interesting question of the role of structural representation in the behavior of the CG models. It is noted that the AR3 model more closely recapitulates individual peptide conformations, which may be interpreted in terms of steric effects. More generally, it is likely that increasing the number of constrained degrees of freedom simply produces a smoother approximation of the potential energy surface for CG configurations that closely resemble sampled AA configurations.

Nevertheless, these results demonstrate that the behavior of CG models is sensitive to structural representation and that different representations may be useful in discovering correspondingly different mechanisms of aggregation. Using the hybrid approach described here, an experimenter may benchmark a relatively simple representation of the target molecule and produce a contrasting model to uncover elements of chemical behavior that the earlier representation may have missed. The judicious introduction of additional degrees of freedom allows one to resolve chemical details in CG models of aromatic rings, but these improvements are not limited to phenylalanine or rings either. Interesting moieties, including anisotropic or rigid molecular structures, can be targeted for finer-grained CG representation, therefore, enabling a more exact analysis of their influence on the mechanisms underlying aggregation.

The structure- and force-based methodologies provide a convenient and rigorous platform for the *ad hoc* creation of CG models of a broad range of soft matter systems. The loss of structure in the backmapped simulations implies a mismatch between the CG model and the AA model it approximates, which diminishes direct comparison between the simulated nanostructures and those found previously in experiments. Conversely, backmapped simulations may in future provide a useful feedback mechanism in the bottom-up model development process.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00458.

Details of molecular representations in each model. Tables of simulated system concentrations and simulation times. Plots of bonded parameters, interparticle RDFs, end-to-end distances, and R_g for all simulated concentrations. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Meenakshi Dutt − *Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0001-5383-2992; Email: meenakshi.dutt@rutgers.edu

Authors

Mason Hooten − *Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States*

Akash Banerjee − *Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-1192-2727

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00458

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.D. acknowledges financial support from NSF CAREER DMR-1654325, DMREF-2118860, and OAC-1835449 awards. The authors acknowledge the use of computational resources enabled via an allocation from NSF ACCESS (allocation DMR-140125).

■ REFERENCES

(1) Stanković, I. M.; Niu, S.; Hall, M. B.; Zarić, S. D. Role of Aromatic Amino Acids in Amyloid Self-Assembly. *Int. J. Biol. Macromol.* **2020**, *156*, 949−959.

(2) Wang, J.; Liu, K.; Xing, R.; Yan, X. Peptide Self-Assembly: Thermodynamics and Kinetics. *Chem. Soc. Rev.* **2016**, *45* (20), 5589−5604.

(3) Hashimoto, M.; Ho, G.; Takamatsu, Y.; Wada, R.; Sugama, S.; Takenouchi, T.; Waragai, M.; Masliah, E. Possible Role of Amyloid Cross-Seeding in Evolvability and Neurodegenerative Disease. *Journal of Parkinson's Disease* **2019**, *9* (4), 793−802.

(4) Panda, J. J.; Chauhan, V. S. Short Peptide Based Self-Assembled Nanostructures: Implications in Drug Delivery and Tissue Engineering. *Polym. Chem.* **2014**, *5* (15), 4431−4449.

(5) Mushnoori, S.; Lu, C. Y.; Schmidt, K.; Zang, E.; Dutt, M. Peptide-Based Vesicles and Droplets: A Review. *J. Phys.: Condens. Matter* **2020**, *33* (5), 053002.

(6) Mishra, A.; Panda, J. J.; Basu, A.; Chauhan, V. S. Nanovesicles Based on Self-Assembly of Conformationally Constrained Aromatic Residue Containing Amphiphilic Dipeptides. *Langmuir* **2008**, *24* (9), 4571−4576.
(7) Scognamiglio, P. L.; Riccardi, C.; Palumbo, R.; Gale, T. F.; Musumeci, D.; Roviello, G. N. Self-Assembly of Thyminyl l-Tryptophanamide (TrpT) Building Blocks for the Potential Development of Drug Delivery Nanosystems. *J. Nanostruct Chem.* **2023**, DOI: 10.1007/s40097-023-00523-7.
(8) Reches, M.; Gazit, E. Formation of Closed-Cage Nanostructures by Self-Assembly of Aromatic Dipeptides. *Nano Lett.* **2004**, *4* (4), 581−585.
(9) Panda, S. S.; Katz, H. E.; Tovar, J. D. Solid-State Electrical Applications of Protein and Peptide Based Nanomaterials. *Chem. Soc. Rev.* **2018**, *47* (10), 3640−3658.
(10) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. Self-Assembly of Phenylalanine Oligopeptides: Insights from Experiments and Simulations. *Biophys. J.* **2009**, *96* (12), 5020−5029.
(11) Mayans, E.; Casanovas, J.; Gil, A. M.; Jiménez, A. I.; Cativiela, C.; Puiggalí, J.; Alemán, C. Diversity and Hierarchy in Supramolecular Assemblies of Triphenylalanine: From Laminated Helical Ribbons to Toroids. *Langmuir* **2017**, *33* (16), 4036−4048.
(12) Xiong, Q.; Liu, Z.; Han, W. Sequence-Dependent Nanofiber Structures of Phenylalanine and Isoleucine Tripeptides. *International Journal of Molecular Sciences* **2020**, *21* (22), 8431.
(13) Han, T. H.; Ok, T.; Kim, J.; Shin, D. O.; Ihee, H.; Lee, H.-S.; Kim, S. O. Bionanosphere Lithography via Hierarchical Peptide Self-Assembly of Aromatic Triphenylalanine. *Small* **2010**, *6* (8), 945−951.
(14) Mayans, E.; Ballano, G.; Casanovas, J.; del Valle, L. J.; Perez-Madrigal, M. M.; Estrany, F.; Jimenez, A. I.; Puiggali, J.; Cativiela, C.; Aleman, C. Hierarchical Self-Assembly of Di-, Tri- and Tetraphenylalanine Peptides Capped with Two Fluorenyl Functionalities: From Polymorphs to Dendrites. *Soft Matter* **2016**, *12* (24), 5475−5488.
(15) Pérez-Madrigal, M. M.; Gil, A. M.; Casanovas, J.; Jiménez, A. I.; Macor, L. P.; Alemán, C. Self-Assembly Pathways in a Triphenylalanine Peptide Capped with Aromatic Groups. *Colloids Surf., B* **2022**, *216*, 112522.
(16) Ozawa, Y.; Sato, H.; Kayano, Y.; Yamaki, N.; Izato, Y.; Miyake, A.; Naito, A.; Kawamura, I. Self-Assembly of Tripeptides into γ-Turn Nanostructures. *Phys. Chem. Chem. Phys.* **2019**, *21* (21), 10879−10883.
(17) Kubota, R.; Torigoe, S.; Liu, S.; Hamachi, I. In Situ Real-Time Confocal Imaging of a Self-Assembling Peptide-Grafted Polymer Showing pH-Responsive Hydrogelation. *Chem. Lett.* **2020**, *49* (11), 1319−1323.
(18) Banerjee, A.; Dutt, M. A Hybrid Approach for Coarse-Graining Helical Peptoids: Solvation, Secondary Structure, and Assembly. *J. Chem. Phys.* **2023**, *158* (11), 114105.
(19) Banerjee, A.; Dutt, M. Self-Organization of Mobile, Polyelectrolytic Dendrons on Stable, Amphiphile-Based Spherical Surfaces. *Langmuir* **2023**, *39* (9), 3439−3449.
(20) Banerjee, A.; Lu, C. Y.; Dutt, M. A Hybrid Coarse-Grained Model for Structure, Solvation and Assembly of Lipid-like Peptides. *Phys. Chem. Chem. Phys.* **2022**, *24*, 1553.
(21) Mushnoori, S.; Schmidt, K.; Nanda, V.; Dutt, M. Designing Phenylalanine-Based Hybrid Biological Materials: Controlling Morphology via Molecular Composition. *Org. Biomol. Chem.* **2018**, *16* (14), 2499−2507.
(22) Aydin, F.; Dutt, M. Surface Reconfiguration of Binary Lipid Vesicles via Electrostatically Induced Nanoparticle Adsorption. *J. Phys. Chem. B* **2016**, *120* (27), 6646−6656.
(23) Aydin, F.; Ludford, P.; Dutt, M. Phase Segregation in Bio-Inspired Multi-Component Vesicles Encompassing Double Tail Phospholipid Species. *Soft Matter* **2014**, *10* (32), 6096−6108.
(24) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Triphenylalanine Peptides Self-Assemble into Nanospheres and Nanorods That Are Different from the Nanovesicles and Nanotubes Formed by Diphenylalanine Peptides. *Nanoscale* **2014**, *6* (5), 2800−2811.
(25) Noid, W. G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139* (9), 090901.
(26) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodriguez-Ropero, F.; van der Vegt, N. F. A. Systematic Coarse-Graining Methods for Soft Matter Simulations − a Review. *Soft Matter* **2013**, *9* (7), 2108−2119.
(27) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving Effective Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24* (13), 1624−1636.
(28) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109* (7), 2469−2473.
(29) Shell, M. S. The Relative Entropy Is Fundamental to Multiscale and Inverse Thermodynamic Problems. *J. Chem. Phys.* **2008**, *129* (14), 144108.
(30) Villa, A.; Peter, C.; van der Vegt, N. F. A. Self-Assembling Dipeptides: Conformational Sampling in Solvent -Free Coarse-Grained Simulation. *Phys. Chem. Chem. Phys.* **2009**, *11* (12), 2077−2086.
(31) Villa, A.; van der Vegt, N. F. A.; Peter, C. Self-Assembling Dipeptides: Including Solvent Degrees of Freedom in a Coarse-Grained Model. *Phys. Chem. Chem. Phys.* **2009**, *11* (12), 2068−2076.
(32) Carmichael, S. P.; Shell, M. S. A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly. *J. Phys. Chem. B* **2012**, *116* (29), 8383−8393.
(33) Ozgur, B.; Sayar, M. Representation of the Conformational Ensemble of Peptides in Coarse Grained Simulations. *J. Chem. Phys.* **2020**, *153* (5). DOI: 10.1063/5.0012391.
(34) Hills, R. D., Jr; Lu, L.; Voth, G. A. Multiscale Coarse-Graining of the Protein Energy Landscape. *PLOS Computational Biology* **2010**, *6* (6), No. e1000827.
(35) Hooten, M.; Dutt, M. Hybrid Bottom-Up Coarse-Grained Model for Aromatic Peptides, 2023. *GitHub repository*. https://github.com/duttm/Hybrid_Bottom-Up_Coarse-Grained_Model_for_Aromatic_Peptides (accessed 2023-04-28).
(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(37) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049−1074.
(38) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269−6271.
(39) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.
(41) 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*, 19−25.
(42) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43−56.
(43) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model* **2001**, *7* (8), 306−317.
(44) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(45) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach. *Biophys. J.* **2007**, *92* (12), 4289−4303.
(46) Rühle, V.; Junghans, C. Hybrid Approaches to Coarse-Graining Using the VOTCA Package: Liquid Hexane. *Macromol. Theory Simul.* **2011**, *20* (7), 472−477.

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(47) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *J. Chem. Theory Comput.* **2009**, *5* (12), 3211−3223.
(48) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, Second ed..; Oxford University Press: Oxford, United Kingdom, 2017.
(49) Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P. Peter; Groningen Biomolecular Sciences and Biotechnology; Molecular Dynamics. *Going Backward. Journal of Chemical Theory and Computation* **2014**, *10* (2), 676−690.
(50) Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F. A.; Kremer, K. Multiscale Modeling of Soft Matter: Scaling of Dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10412−10420.
(51) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42* (16), 6801−6822.
(52) Brown, S. T.; Buitrago, P.; Hanna, E.; Sanielevici, S.; Scibek, R.; Nystrom, N. A. Bridges-2: A Platform for Rapidly-Evolving and Data Intensive Research. *Practice and Experience in Advanced Research Computing*; PEARC '21; Association for Computing Machinery: New York, NY, USA, 2021; pp 1−4. DOI: 10.1145/3437359.3465593.
(53) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16* (5), 62−74. http://doi.ieeecomputersociety.org/10.1109/MCSE.2014.80
(54) 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* (23), 234115.

https://doi.org/10.1021/acs.jctc.3c00458
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX