

Development, Validation, and Applications of Nonbonded Interaction Parameters between Coarse-Grained Amino Acid and Water Models

Esmat Mohammadi, Soumil Y. Joshi, and Sanket A. Deshmukh*



Cite This: *Biomacromolecules* 2023, 24, 4078–4092



Read Online

ACCESS |



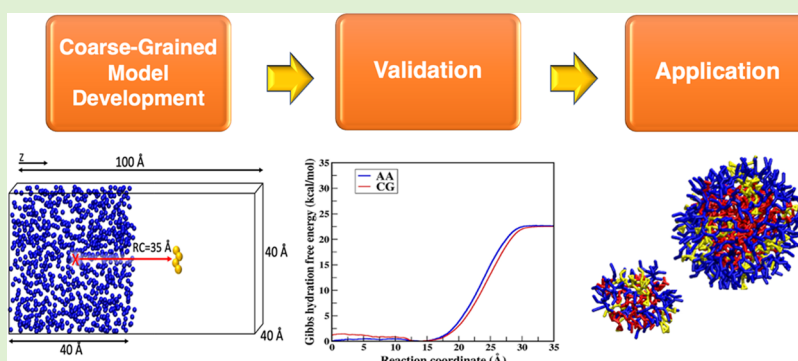
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Interactions between amino acids and water play an important role in determining the stability and folding/unfolding, in aqueous solution, of many biological macromolecules, which affects their function. Thus, understanding the molecular-level interactions between water and amino acids is crucial to tune their function in aqueous solutions. Herein, we have developed nonbonded interaction parameters between the coarse-grained (CG) models of 20 amino acids and the one-site CG water model. The nonbonded parameters, represented using the 12–6 Lennard Jones (LJ) potential form, have been optimized using an artificial neural network (ANN)-assisted particle swarm optimization (PSO) (ANN-assisted PSO) method. All-atom (AA) molecular dynamics (MD) simulations of dipeptides in TIP3P water molecules were performed to calculate the Gibbs hydration free energies. The nonbonded force-field (FF) parameters between CG amino acids and the one-site CG water model were developed to accurately reproduce these energies. Furthermore, to test the transferability of these newly developed parameters, we calculated the hydration free energies of the analogues of the amino acid side chains, which showed good agreement with reported experimental data. Additionally, we show the applicability of these models by performing self-assembly simulations of peptide amphiphiles. Overall, these models are transferable and can be used to study the self-assembly of various biomaterials and biomolecules to develop a mechanistic understanding of these processes.

1. INTRODUCTION

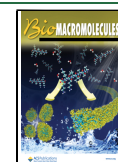
Among the hundreds of amino acids found in nature, 20 amino acids make up all proteins in the human body.^{1–3} The interactions of these amino acids with water are instrumental for accurate protein folding and maintaining the stability of these complex structures, thus influencing their overall functions.^{4–6} For example, hydrophobic amino acids, when clustered in the interior of cytoplasmic proteins, may form ligand-binding pockets essential for signaling and enzyme catalysis.^{7–9} Similarly, hydrophilic amino acids in the interiors of transmembrane proteins are essential for extracellular transport of water and ions, thus maintaining cellular homeostasis.^{10,11} Moreover, interactions between amino acids and water are advantageous in creating novel peptide-based (macro)molecules including artificial peptide polymers and peptide amphiphiles with unique properties and function-

alities.^{12–14} A combination of both hydrophilic and hydrophobic amino acids in elastin-like polypeptides (ELPs)—a class of artificial peptide polymers—and their interactions with water is believed to be responsible for their lower critical solution temperature (LCST) behavior.^{13,15} In the case of peptide amphiphiles (PAs), the sequence of peptides, their concentration, and interactions with water results in self-assembled structures, such as micelles, fibers, rods, and many more, regulated by environmental conditions such as temper-

Received: April 28, 2023

Revised: July 25, 2023

Published: August 21, 2023



ature and pH.^{16–19} Several experimental and computational studies have shown the self-assembly of PAs into micelles and consequently fibers, having micron-sized lengths and nanometer-sized diameters (1–10 nm).^{12,18,20–22} Moreover, self-assembled structures of PAs have shown great promise in biomedical applications like bone regeneration, tissue engineering, gene and drug release, regenerative medicine, etc.^{18,22,23} Hence, understanding amino acid–water interactions is essential in order to gain insight into the conformations and function of many biomolecules in aqueous solutions.

Processes such as self-assembly and folding/unfolding of proteins occur at the length and time scales of nanometers and microseconds, respectively.^{24–26} Although the development of novel experimental methods has enabled accurate characterization of these self-assembled nanostructures, following the self-assembly pathway still remains challenging.²⁷ Consequently, during the past decades, computational methods such as coarse-grained (CG) molecular dynamics (MD) simulations have drawn a lot of attention to provide in-depth insights into the mechanisms governing these molecular processes.^{12,14,28–32} However, the accuracy of these results depends significantly on the intra- and intermolecular interaction parameters between the CG beads and the solvent, defined by the force-field (FF).^{24–26} Several research groups have developed CG models of amino acids, but most of these studies have used implicit water models, which makes it impossible to understand the structure of the solvent at the solute–solvent interfaces.³³ Of the few models that can be used with explicit water beads,^{34–36} the Martini FF, one of the most common models for protein CG MD simulations, is also known to have limitations like its inability to predict the Gibbs hydration free energy of CG molecules.^{24,37–39} In addition, these models may underestimate the characteristics of self-assembled structures and may not be suitable to study elastin-like peptides that exhibit a lower critical solution temperature (LCST).^{12,40,41} Thus, it is important to develop FF parameters to accurately model the interactions between amino acids and water. Han et al. developed a CG protein model by employing a united atom representation of all 20 amino acids.⁴² The nonbonded interactions were tuned by fitting experimentally obtained properties such as densities, self-solvation free energies, and hydration free energies of over 100 organic molecules, and the transfer free energies of side-chain analogues of various amino acids from cyclohexane to water were used to assess the model's performance. Similarly, Ha-Duong et al. introduced a numerically efficient CG water model to simulate the solvation effects of CG protein models developed by Basdevant et al.^{43,44} Their water model was represented using polarizable pseudo-particles and accurately described solvation properties, such as hydrophobic forces between two hydrophobic solutes (amino acids) and electrostatic solvation free energies of 17 peptides, which were further validated by comparison with all-atom MD simulation results. However, these amino acid models in the CG water were not tested for the solvation of polypeptides or investigated for other experimental behaviors (e.g., self-assembly of peptide amphiphiles). Furthermore, additional studies on CG protein models can be found in several review papers.^{45–51}

In the present study, we employed an intermediate mapping scheme for amino acids, balancing between the united atom model and larger CG beads used in the existing models.^{46,52–55} This scheme retains the overall shape and local structure (e.g., rings, hydrophilic/hydrophobic groups, etc.) of the amino

acids while facilitating computationally efficient MD simulations of larger systems for up to several microseconds. The FF parameters for these structurally accurate CG models of amino acids, which can capture several experimental dipeptide properties, including density, surface tension, and heat of vaporization, have been previously developed in our group.⁵⁶ Herein, using these transferable CG amino acid models, we developed the nonbonded interaction parameters between amino acids and our 2:1 mapped water model such that they can capture the Gibbs hydration free energy of dipeptides obtained from all-atom (AA) MD simulations. Note, our CG water model has demonstrated accurate reproduction of various important physical, structural, dynamical, and thermodynamic properties of water, and it has been used with different soft-material systems.^{24,25,33,57–59} In the present study, to validate the accuracy and test the transferability of the newly developed interaction parameters between amino acid and water CG beads, we calculated the hydration free energies of the analogues of their side chains, which showed a good agreement for most cases with an error within 10% as compared to their experimental values. We have employed these parameters in CG MD simulations to study the self-assembly of two different PAs, namely, c16-AHL3K3-CO2H and (AF)₆H₅K₁₅ (also known as FA32), thus demonstrating their applications in studying self-assembly processes on biomolecules.¹²

2. METHODS AND COMPUTATIONAL DETAILS

2.1. Force-Field Equations. In order to conduct CG MD simulations, we used CG mapping schemes that were previously developed in our group, including a one-site water model representing two water molecules in one CG bead⁵⁹ and CG amino acids possessing 2:1, 3:1, and 4:1 mapping schemes at most (having 2, 3, and 4 heavy atoms in one bead with their associated hydrogens).⁵⁶ The mapping schemes of the 20 amino acids used in this study are shown in Table S1. Note that all of the CG beads were parameterized to be charge-neutral in order to accelerate MD simulations by eliminating the time-consuming electrostatic energy calculations. Hence, similar to many existing CG models, they cannot be used to study processes driven by pH and/or where pH plays a critical role.^{45–47,49,50,60} Moreover, charge neutrality also limits their applications in investigating systems where electrostatic interactions are critical. Consequently, caution must be exercised when using these models to ensure that they are suitable to solve the problem at hand.

The FF equation used to define the interactions between CG beads is presented below

$$U = \sum k_b(b_0 - b)^2 + \sum k_\theta(\Theta_0 - \Theta)^2 + \sum k_\phi(1 + \cos(n\phi - \delta)) + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1)$$

Here, the first three terms represent energy contributions from bonded interactions between CG beads—including bonds, angles, and dihedrals. The nonbonded self- and cross-interactions are represented through the 12–6 Lennard Jones (LJ) potential as shown in the last term of eq 1. The variable ϵ_{ij} is the potential well depth, which describes the strength of interaction energy between two beads, σ_{ij} is the finite distance where the potential between two beads is equal to zero, and r_{ij} represents the distance between the centers of two interacting beads.^{61,62} The 12–6 LJ equation was utilized to represent the cross-interactions between CG amino acid and water beads to maintain consistency with their respective self-interaction FFs developed earlier.^{57,59} The Lorentz–Berthelot (LB) combining rule is a common method used to determine the parameters for cross-interactions between different types of beads in molecular

simulations.^{63,64} Using this method, ϵ_{ij} and σ_{ij} between two atoms i and j can be calculated using equations

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2} \quad (2)$$

$$\sigma_{ij} = 1/2(\sigma_{ii} + \sigma_{jj}) \quad (3)$$

These values, calculated for the different interacting bead pairs in our amino acid–water systems, provided a good starting point for further optimization of cross-interactions in our study by reproducing Gibbs hydration free energy values of amino acid dipeptides.

2.2. Computational Details of Gibbs Hydration Free Energy Simulations. Reproducing the solvation free energy has been successfully employed for tuning the solvent–solute interactions in several previous studies.^{24,59,65–67} Gibbs hydration free energies of amino acid dipeptides were calculated through AA adaptive biasing force (ABF) MD simulations using the CHARMM FF and TIP3P water model. The same dipeptides, as utilized in our group previously to establish bonded parameters for CG amino acids,⁵⁶ were chosen to calculate the AA free energies. Chemical structures and the corresponding CG representations for these dipeptides are presented in Table S2. The values obtained from the AA simulations were used as targets to optimize CG amino acid–water interactions.

All MD simulations in this study were conducted using the NAMD 2.14 software and the Colvars package.^{68–70} The ABF simulations were performed in the NVT ensemble, where the temperature was maintained using the Langevin thermostat.⁷¹ A switching function was applied to change the functional form of van der Waals interaction (VDW) at a distance of 9 Å so that the VDW potential energy could be truncated at a cutoff distance equal to 12 Å. A pair list distance was also used at 15 Å, which refers to the distance within which NAMD searches for the atoms that interact through the VDW potential. For AA simulations, we used a time step equal to 1 fs, while for CG simulations, the time step was chosen to be 10 fs. The initial solvation configuration files of the solvation of amino acids in water were generated using the Packmol software.⁷²

A schematic of the simulation setup to perform ABF simulation is illustrated in Figure 1. Initially, one dipeptide molecule of amino acid

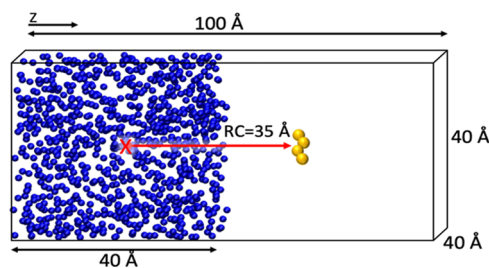


Figure 1. Schematic representation of obtaining Gibbs hydration free energy for an amino acid molecule in a water box. The blue beads illustrate water molecules and the yellow beads represent CG dipeptide molecules. The z -direction, depicted by an arrow, shows the reaction coordinate (RC), which is at a 35 Å distance.

was solvated in the center of a box of water with 40 Å × 40 Å × 40 Å dimensions. This box contained 950 CG water beads (1900 AA water molecules). The air–water interface was created by extending the z -direction to 80 Å. Then, the molecule was transferred from water to vacuum by moving it along the z -axis at a 35 Å distance from the center of mass of the water box. This 35 Å is also referred to as the reaction coordinate (RC), which is a measure of the distance between the center of mass of the water box and that of a CG amino acid molecule in the z -direction. The periodic boundary condition was applied in three directions. The Gibbs hydration free energy profile was obtained along the z -axis from zero (COM of water box) to 35 Å (vacuum). To enhance the efficiency of ABF simulations, the entire RC was also divided into seven consecutive 5 Å wide windows. For each of these windows, a separate MD simulation was conducted with

the simulation run equal to 20 ns at a temperature equal to 300 K. Consequently, the total simulation time for obtaining the hydration free energy of one amino acid dipeptide was equal to 140 ns. Gibbs hydration free energy for each amino acid was obtained from three independent sets of simulations and was averaged out in order to reduce the statistical error.

2.3. ANN-Assisted PSO Algorithm for Parameter Development. For optimizing amino acid–water interactions in our CG models, we calculated σ values between interacting bead pairs using the LB combining rule and optimized the ϵ values between amino acid beads and the one-site water model to reproduce hydration free energies of amino acid dipeptides. Equation 4 was used to calculate the errors between free energies obtained from AA and CG simulations

$$\begin{aligned} \text{error \%} &= \frac{|(\text{CG hydration free energy} - \text{AA hydration free energy})|}{\text{AA hydration free energy}} \\ &\times 100 \end{aligned} \quad (4)$$

It is important to note that the focus of parameter tuning was specifically on developing the interactions between CG beads of amino acids and water models. These interactions were defined using a pair-specific nonbonded LJ parameter defined using the NBFIX (nonbonded fix) method, a commonly used approach in several all-atom force fields.^{73–75} NBFIX allows for the adjustment of nonbonded interactions between specific atom pairs or groups to better reproduce experimental data and to improve the accuracy of simulations.

In order to automate and accelerate the process of optimizing ϵ values, the artificial neural network (ANN)-assisted particle swarm optimization (PSO) algorithm was utilized. The rationale behind employing the ANN-assisted PSO in our parametrization process lies in the optimization of multiple parameters for various amino acids, shown in Table 1, with several parameters being shared between multiple amino acids. Thus, even though we have only used the Gibbs Free energy for each dipeptide as a target property, these large number of parameters (total 36) between CG beads necessitate the utilization of advanced optimization techniques. This is mainly because traditional trial-and-error, manual approaches for parameter development can be time-consuming and challenging. Based on our group's prior experience,^{24,56–59,76–78} we used ANN-assisted PSO to accelerate the parameter development process.

The PSO algorithm, inspired by the motion of a flock of birds searching for food, was employed to generate 31 initial sets of parameters (birds).^{28,56–59,76–78} Note that these initial sets of ϵ values were randomly chosen from the optimization ranges for each of the variables to be optimized. After generating the initial parameters, MD simulations were performed to calculate the resulting free energy values, and the corresponding errors were calculated. This algorithm further iterated the parameter values until the error was within an acceptable threshold (<5%) or the total number of iterations (300) was achieved. More details can also be found in Section S1, and a flowchart demonstrating the working of the PSO algorithm is presented in Figure S1.

Typically, ANN consists of one input layer, one output layer, and multiple hidden layers, with the number of input and output nodes equal to the number of target properties and variables.⁷⁹ Herein, we used a network comprising one node in the input layer (Gibbs hydration free energy), and the number of nodes in the output layer was equal to the total number of ϵ parameters between CG amino acid beads and water for a given amino acid. More details of the ANN model can be found in Section S2, and a schematic of the model is illustrated in Figure S2. Although the PSO algorithm implemented with 31 particles is instrumental in efficiently exploring the parameter space and accelerating the optimization of force-field (FF) parameters, incorporating ANN within the PSO framework offers several advantages. The ANN model not only expedites the optimization process by achieving a satisfactory minimal error sooner than using PSO alone, but also provides guidance in model

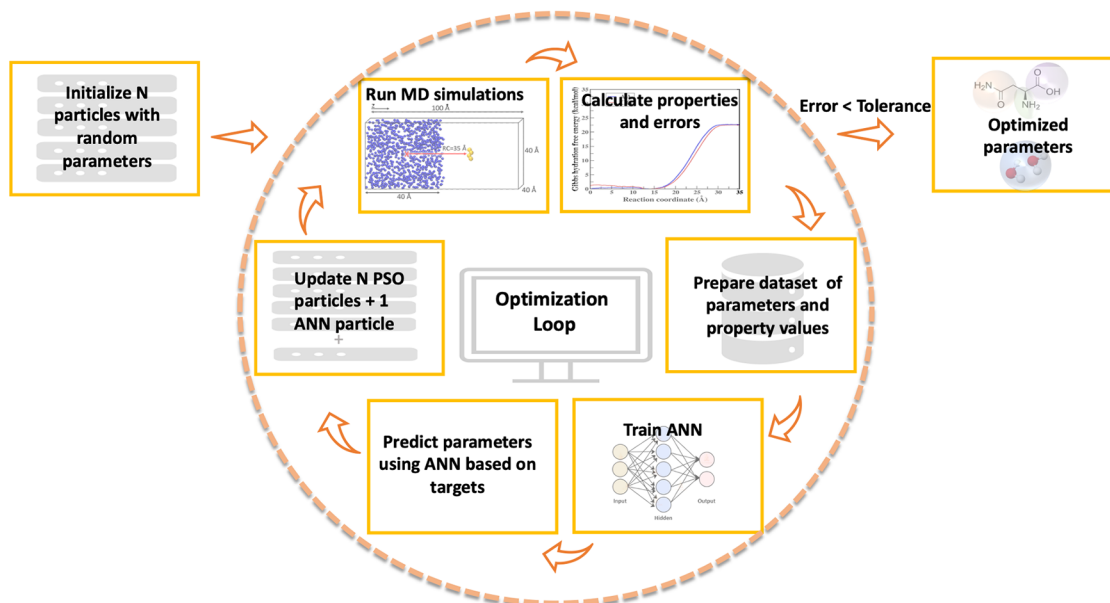


Figure 2. Schematic representation of the ANN-assisted PSO algorithm.

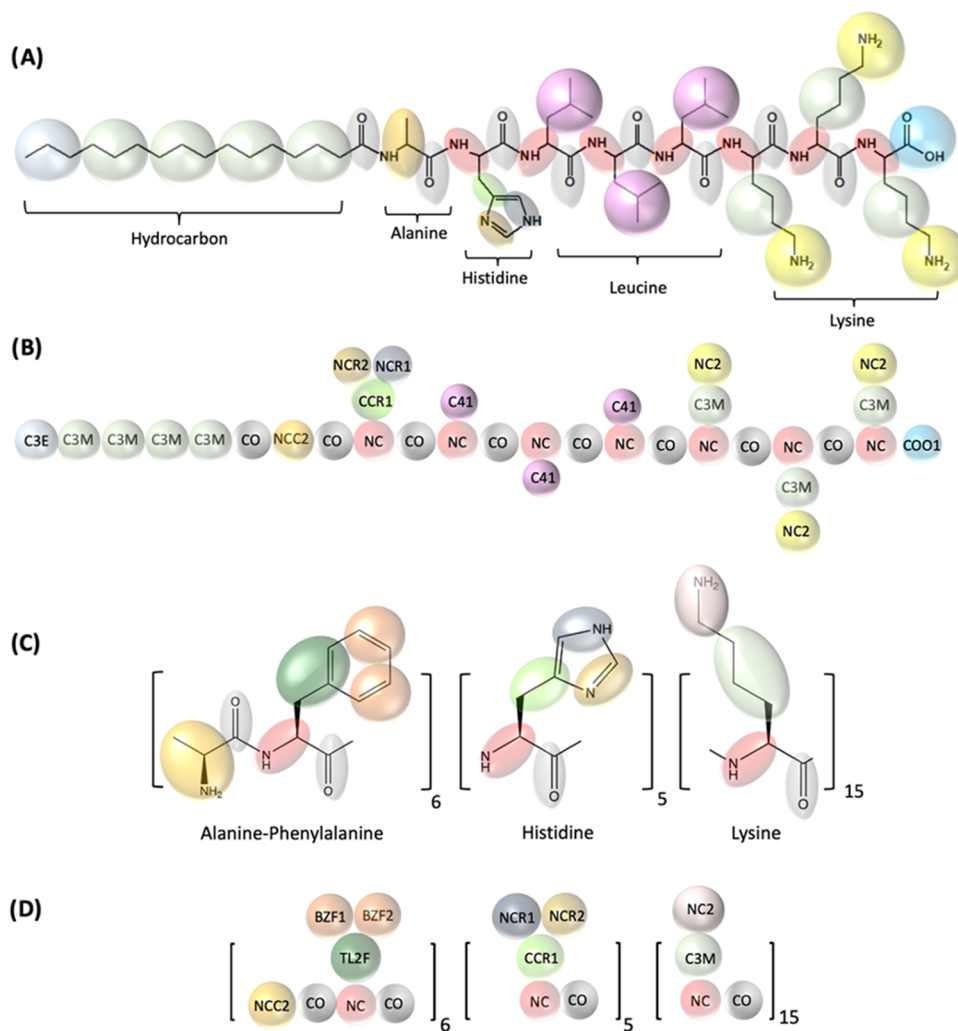


Figure 3. Schematic AA and CG mapping structures of c16-AHL3K3-CO₂H (A, B) and (AF)6H5K15 (C, D), respectively.

Table 2. Free Energy Values of Analogues of Amino Acid Side Chains Calculated from CG Models Using Optimized ϵ Values along with the Experimental Values Reported in the Literature^{91–96}

amino acid	side-chain analogues	experimental hydration free energy of analogues (kcal/mol)	hydration free energy of analogues from MD (kcal/mol)	error
alanine	<i>N</i> -butylamine	−4.30	−4.67 ± 0.16	8.46
asparagine	acetamide	−9.68	−8.73 ± 0.11	9.86
aspartic acid	acetic acid	−6.486	−6.29 ± 0.17	2.95
cysteine	methanethiol	−1.24	0.43 ± 0.33	65.81
glutamine	propionamide	−9.38	−8.66 ± 0.72	7.63
phenylalanine	toluene	−0.76	−0.72 ± 0.18	5.18
serine	methanol	−5.06	−4.81 ± 0.042	4.98
threonine	ethanol	−4.88	−3.40 ± 0.19	30.29
tyrosine	<i>m</i> -cresol	−6.11	−5.51 ± 0.08	9.78
valine	2,3-dimethylbutane	4.20	1.85 ± 0.16	55.99
glutamic acid, C-terminus	propionic acid	−6.36	−5.81 ± 0.23	8.64
lysine	butylamine	−4.30	−7.06 ± 0.16	64.09
histidine	pyridine	−4.7	−4.72 ± 0.14	0.36
	pyrrole	−4.79	−4.64 ± 0.12	3.09
	toluene	−0.76	−0.73 ± 0.24	4.25
tryptophan	pyrrole	−4.79	−4.36 ± 0.10	8.99
	toluene	−0.76	−0.79 ± 0.16	4.13
glycine, N-terminus	<i>N</i> -butylamine	−4.30	−4.29 ± 0.39	0.23

these analogues, we believe that they served as a valuable means to assess the accuracy and transferability of the newly developed parameters. The approach mentioned in Section 2.2 was used to conduct these free energy simulations and the results were compared to published experimental values.

2.5. Applications of FF Parameters to Study the Self-Assembly of Peptide Amphiphiles. The developed FF parameters between amino acids and water were further used to perform MD simulations investigating the self-assembly behaviors of two PA systems, namely, c16-AHL3K3-CO₂H and FA32.^{12,81–83} The PA c16-AHL3K3-CO₂H includes a hydrocarbon tail, represented by C3E and C3M CG beads that were developed in our research group.²⁴ These hydrocarbon beads interact with water molecules through a 12–6 LJ potential, which was fine-tuned to accurately reproduce the experimentally measured Gibbs hydration free energies of various hydrocarbons. The parameters for the hydrocarbon beads, including their ϵ and σ values are listed in Table S4. The hydrophobic tail is connected to a polypeptide chain made up of four different amino acids: alanine, histidine, leucine, and lysine. We mapped the peptide chain of the molecule using our CG models and utilized the tuned CG parameters for the interaction between the amino acids and water. Because the backbone of all four amino acids was represented by the same CG beads (NC and CO), we were able to use our CG mapping to simulate the combination of these peptides. Similarly, FA32 consists of 32 amino acids, including alanine, phenylalanine, histidine, and lysine. Due to their hydrophobic nature, alanine and phenylalanine are expected to form the core of the self-assembled structure, while lysine, a hydrophilic amino acid, is expected to be present at the interface between the self-assembled structure and water. The CG mapping schemes for c16-AHL3K3-CO₂H and FA32 are shown in Figure 3.^{12,81–85}

In the case of c16-AHL3K3-CO₂H simulations, 150 CG PAs were randomly and homogeneously dispersed in an initial box size of 87 Å × 114 Å × 185 Å containing 27,850 one-site water beads. The CG MD simulations were conducted for 5 μ s in an NPT ensemble at 340 K temperature. On the other hand, for the simulations of FA32, nine different concentrations were used because it is known that the concentration of peptides is a crucial determinant in governing the process of assembly. We used 12, 18, 24, 30, 36, 42, 48, 54, and 60 PA molecules inside a box of 180 Å × 180 Å × 180 Å containing 85,000 one-site water beads. Employing our amino acid models, a relaxed single chain was represented using 127 beads, as compared to 583 all-atoms. Each simulation was performed at 310 K in the NPT ensemble

for 3.5 μ s. In both self-assembly simulations, the pressure was kept constant at 1 atm, and Langevin thermostat and barostat were used for temperature and pressure control, respectively.⁷⁰ The simulations were stable with a time step of 5 fs and periodic boundary conditions were applied in all three dimensions to simulate an infinite system. The cutoff distance used to truncate the VDW interactions was 12 Å,⁸⁶ and “exclude 1–2” was used for nonbonded interactions.^{82,83}

3. RESULTS AND DISCUSSION

3.1. Optimized Parameters between CG Amino Acid and Water Beads. Prior to the PSO optimization, the LB combining rule was employed to obtain σ and ϵ values between various amino acids and water beads, as shown in Table 1. Figure S3 shows the comparison between AA and CG free energies obtained from the LB combining rule for representative amino acids, namely, asparagine, aspartic acid, cysteine, glutamic acid, glycine, and phenylalanine. As can be seen, the values of Gibbs hydration free energies of amino acid dipeptides (tabulated in Table S4), obtained from these preliminary CG simulations, did not show good agreement with the AA values (minimum error = 25% and maximum error = 136%). This suggests that the ϵ and σ parameters between the amino acids and the water beads, obtained from the LB rule, could not reproduce the target free energy values obtained from AA MD simulations. This is consistent with several previous studies that have reported the failure of the LB rule to capture the experimental properties of gas mixtures and hydrocarbons.^{24,87,88} Therefore, in the present study, it was necessary to optimize the parameters using the ANN-assisted PSO algorithm to improve the accuracy of the simulated values. Keeping σ values equal to those obtained from the LB mixing rule, the optimization ranges of ϵ values were chosen to be approximately $\pm 25\%$ of those obtained by the LB combining rule.

The optimized ϵ values obtained using the ANN-assisted PSO method, as well as σ values calculated using the LB mixing rule, are tabulated in Table 1. The parameter file is also provided in the SI. The Gibbs free energy for 20 dipeptide amino acids obtained from CG MD simulations using the

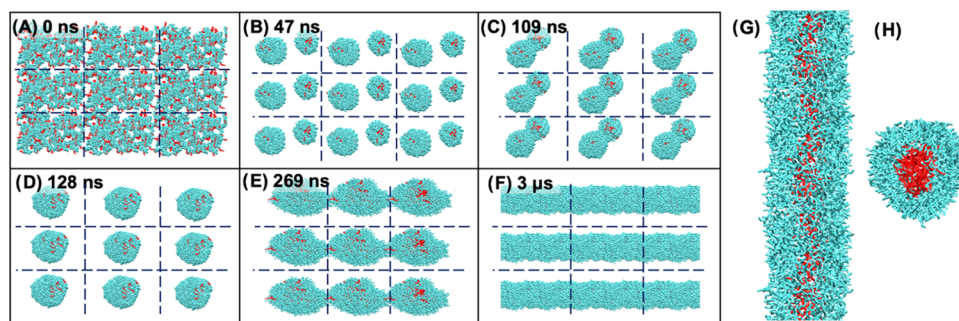


Figure 4. Snapshots of the structural evolution at different time steps from 0 to 3 μ s and the final self-assembled fiber. Snapshot (A) shows the randomly dispersed PAs. Snapshots (B, C) suggest the formation and merging of two separate micelles that aggregate together to form one larger micelle as illustrated in snapshot (D). Snapshot (E) shows the large micelle interacting with its periodic image, resulting in the formation of a fiber (illustrated in snapshot (F)). Dashed lines illustrate the periodic boundaries. Hydrophilic amino acids are shown in cyan, and the hydrophobic tail of PA is shown in red. The CG water molecules are not shown for clarity. Snapshots of the side and front view of the final self-assembled fiber are shown in snapshots (G, H) after 5 μ s.

optimized ϵ parameters along with the results from AA MD simulations is also shown in Table 1. Figure S4 also illustrates the comparison between the free energy profiles of AA and CG simulations. To reduce statistical error, we conducted three independent AA and CG MD simulations, with different initial configurations, for each of the amino acids (total simulation time 15.960 μ s). Using the optimized ϵ values, the results demonstrated a remarkable consistency between the CG and AA free energies, with errors of less than 5%.

As shown in Table 1, the optimized parameters for each amino acid were able to capture the essential features of their respective side chains, such as hydrophobicity and hydrophilicity. For instance, the magnitude of optimized ϵ parameters for hydrophobic amino acids such as phenylalanine (Phe), isoleucine (Ile), leucine (Leu), valine (Val), and proline (Pro) beads are relatively lower than those of hydrophilic amino acids such as serine (Ser), asparagine (Asn), glutamine (Gln), and threonine (Thr). These smaller magnitudes of ϵ values indicate a higher degree of hydrophobicity for the aforementioned hydrophobic amino acids.^{89,90} Conversely, the larger magnitudes of the optimized parameters for hydrophilic amino acids pointed to strong hydrophilicity.^{89,90} Overall, the use of dipeptides during the optimization process allowed us to tailor the CG model to accurately represent the unique features of each amino acid, including their side-chain characteristics.

3.2. Chemical Transferability of New FF Parameters.

To further validate and evaluate the transferability of the newly developed ϵ and σ parameters, the Gibbs hydration free energies of the analogues of the amino acid side chains were calculated using the CG models and compared with the experimental values reported in the literature.^{91–96} The mapping scheme for these analogues is shown in Table S3 and their free energy values are shown in Table 2. Remarkably, for the majority of analogues (14 out of 18), the error between the experimental and simulated values was less than 10%. These findings demonstrate the chemical transferability of the new FF parameters for use with other molecules. The analogues of cysteine, threonine, valine, and lysine side chains, namely, methanethiol, ethanol, 2,3-dimethylbutane, and butylamine, respectively, exhibited higher errors. Thus, for these four analogues, the ϵ and σ parameters between amino acid and water did not show acceptable transferability and further refinement of the parameters may be necessary to improve the accuracy of their simulations. It is worth noting that while most of the MD simulations were conducted at 300 K, simulations

for some, including acetamide, methanethiol, and propionamide, were performed at different temperatures (358, 278, and 356 K, respectively) due to their nonliquid nature at 300 K. These findings suggest that the optimized CG parameters demonstrate good chemical transferability and may be used for other molecules.

3.3. Application of Newly Developed FFs to Study the Self-Assembly of Peptide Amphiphiles.

To test the applicability of the new optimized nonbonded parameters between amino acids and water, we performed the self-assembly of two PAs in water. Specifically, we selected two different PAs, namely, c16-AHL3K3-CO₂H¹⁴ and FA32, that form fibers and micelles, respectively, in the presence of water.^{12,81–85} In general, interactions between water and the hydrophilic and hydrophobic portions of the PAs are known to play an important role in the self-assembly process.

3.3.1. Self-Assembly of c16-AHL3K3-CO₂H. Our group has performed CG MD simulations of the self-assembly of c16-AHL3K3-CO₂H PA molecules using the amino acid and water interaction parameters obtained with LB combining rule for 15 μ s at 340 K.⁹⁷ Experimentally, these PAs are known to self-assemble to form nanofibers.¹² However, in CG MD simulations using LB combining rule, no such nanofibers were observed. Instead, the PAs quickly formed micelle-like structures and bundled together into windowpane structures.⁹⁷ These dynamic and unstable assemblies persisted throughout the simulation, continually breaking apart and regrouping, clearly demonstrating the need to improve the interaction parameters between amino acid and water models to obtain experimentally observed nanofibers.

Herein, we employed our newly developed nonbonded parameters in the PA self-assembly simulation. As illustrated by the snapshots of the structural evolution at different time steps in Figure 4 and a movie provided in SI (c16-AHL3K3-CO₂H.avi), we showed that the randomly dispersed PAs aggregated together in the early stages of the self-assembly and later formed micelles. We observed that after $\sim$ 47 ns, two separate micelles were formed, which got closer to each other in the z -direction. Finally, after $\sim$ 128 ns of the simulation time, the two micelles were in close proximity to each other and aggregated to form one larger micelle at $\sim$ 269 ns. Since we used periodic boundary conditions, the large micelle interacted with its periodic image and formed a perfect fiber in the x -direction, which was found to stabilize after $\sim$ 985 ns. This self-assembly process of micelle, and subsequently, fiber formation

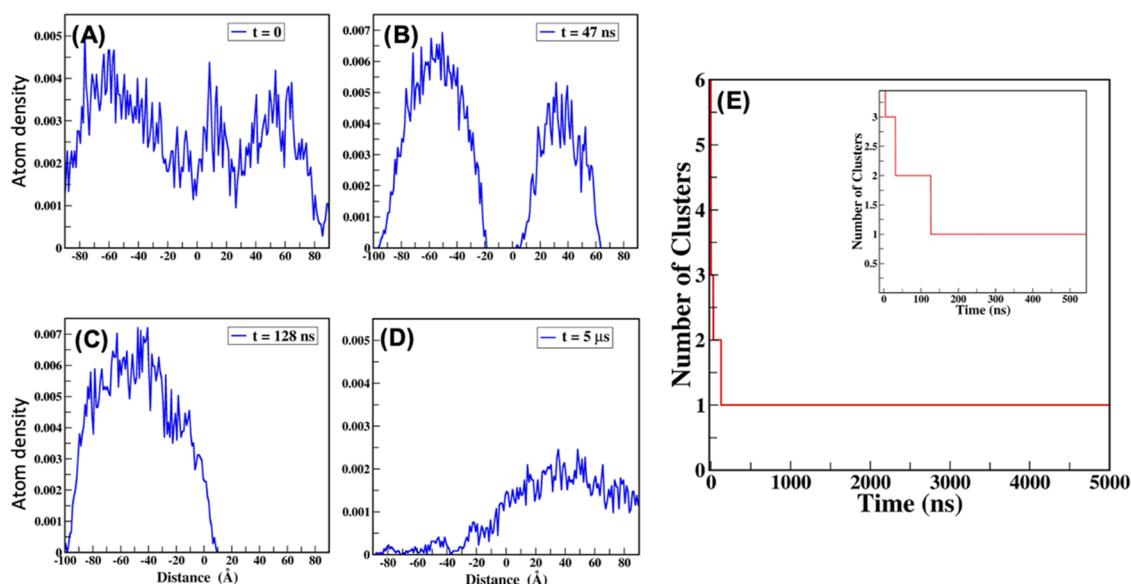


Figure 5. Bead density profiles for PA molecules at various time steps, illustrating the evolution of the self-assembly process (A–D) and the number of clusters of PA molecules versus time for the self-assembly process of c16-AHL3K3-CO2H (E).

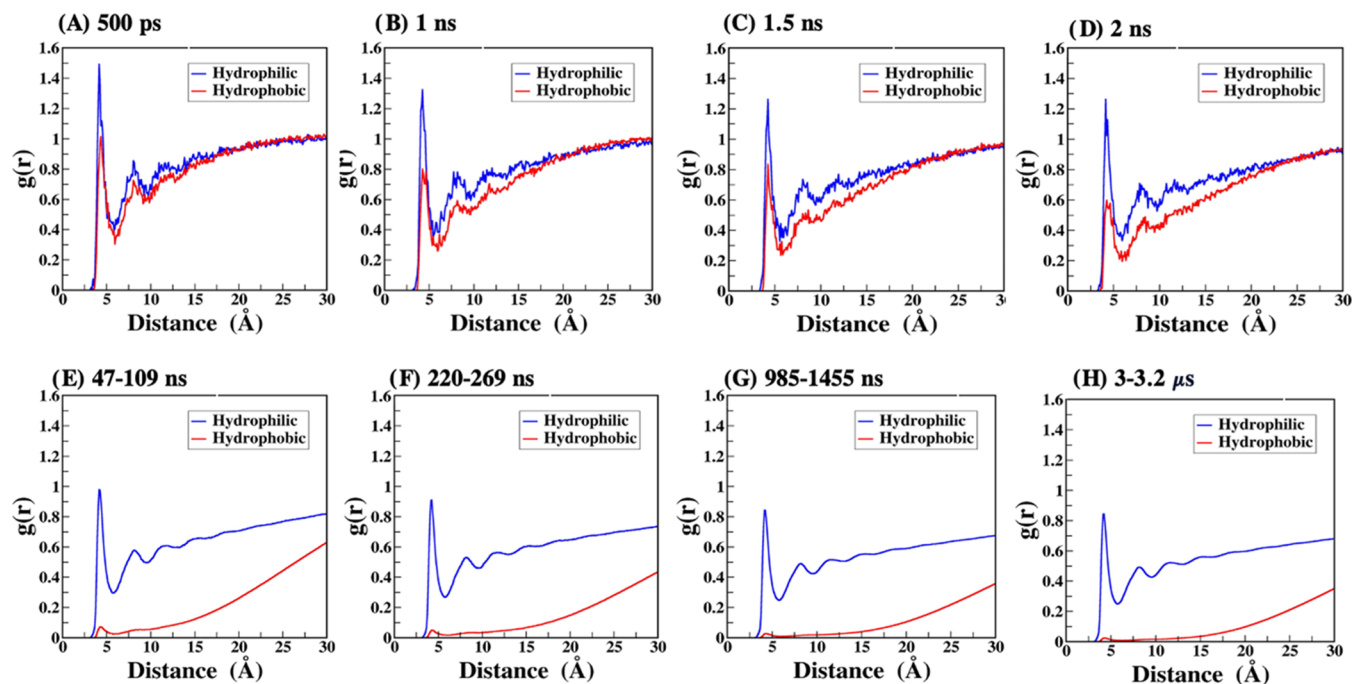


Figure 6. RDF of hydrophilic (lysine) and hydrophobic (hydrocarbon chain) regions of PA at various time steps.

has been previously reported in several experimental and computational studies.^{12,98–102} We monitored various stages in the self-assembly process, including micelle and fiber formation by examining the structural and solvation behavior of the particles in the system. Figure 4G,H illustrates the snapshots of the side and front view of the assembled fiber after 5 μ s.

Figure 5A–D shows the density profiles of the PA molecules wherein the self-assembly process can be clearly observed. The formation of two separate micelles is shown by distinct broad peaks at ~ 47 ns (Figure 5B). With time, these two peaks merge together at ~ 128 ns, demonstrating the formation of a larger micelle (Figure 5C). Finally, at ~ 985 ns, the micelles from periodic images merge and form a fiber (Figure 5D). Note, since the density profiles do not include periodic images,

an illustration of an extended plateau is not possible. Figure S5 shows the density profiles for the hydrophobic tails of the PAs (C16) at different time steps, exhibiting similar trends, albeit with narrower peaks. We calculated the aggregation number of the two initially formed micelles, which consisted of ~ 90 and ~ 60 PAs—in excellent agreement with experimental and computational data reported in the literature.^{22,103–105}

To quantitatively analyze the PA self-assembly process, we employed the density-based spatial clustering of application with noise (DBSCAN) method to examine the self-assembly process of PA molecules.^{106,107} DBSCAN is a popular machine learning-based data clustering algorithm, widely used for the grouping of closely packed data points and thus can be used to count the clusters of molecules.^{106,107} We defined a cluster to

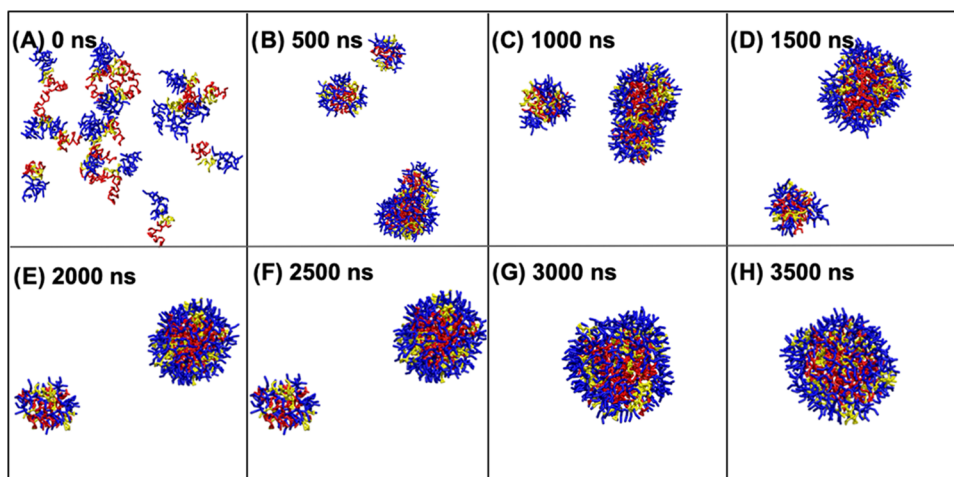


Figure 7. Self-assembly of 18 FA32 polypeptides in a 180 Å water box at different time intervals during the MD simulation. Hydrophobic ala and phe are represented by red beads. Amphiphilic his and hydrophilic lys are shown by yellow and blue beads, respectively. Water molecules are not shown for clarity.

be the aggregation of three or more PAs when the distance between their hydrophobic centers of masses was less than 20 Å. Details of this method can be found in Section S3. As shown in Figure 5E, the number of clusters (N_c) formed in the system decreased as the simulations progressed. During the initial ~50 ns, the number of clusters decreased sharply from six clusters to two, which implies the formation of two micelles. Then, it further decreased to one, showing the formation of one large micelle and subsequently a fiber. This graph was consistent with our other observations as well as the reported literature.¹²

The diameter of self-assembled fibers is an important characteristic that influences their properties and functions.¹⁰⁸ Experimental and computational studies have been previously conducted to measure the diameters of these fibers. In an experimental study on the self-assembly of c16-AHL3K3-CO₂H, it was observed that these molecules self-assembled to form fibers with an approximate diameter of ~69 Å.⁸¹ However, self-assembly simulations carried out using the MARTINI FF have shown the formation of hexagonally packed fibers with ~50 Å diameter.^{12,98} Herein, using our amino acid models, we observed that the self-assembly of c16-AHL3K3-CO₂H molecules resulted in a fiber with a diameter equal to 68.64 ± 0.69 Å, which was consistent with an experimentally reported value of 69 Å.⁸¹ The diameter of our simulated fiber was calculated using an ellipse fitting algorithm that is explained in Section S4 and shown in Figure S7.

Additionally, we examined the solvation behavior of the PAs throughout the simulation time to understand the role of water in the process of self-assembly. The radial distribution functions (RDFs) of both hydrophobic and hydrophilic groups in the PAs with water are illustrated in Figure 6. As shown in this figure, strong structural correlation was observed between water and the hydrophobic alkyl chains at the early stages of self-assembly (<~2 ns). Similar observations have been reported in past studies of self-assembling PAs, which have been attributed to a cage-like structure of water around the hydrophobic tail.^{109–111} During the self-assembly simulations, we observed reduction of RDF peak heights between water and hydrophobic tails, causing the hydrophobic collapse of PAs, driven by the VDW interaction between hydrocarbons. This led to their aggregation and the formation of micelles. We also calculated the end-to-end distances of PAs for 10 randomly

selected molecules in the system. It was observed that the end-to-end distance decreased from ~29 to ~23 Å through the initial stages of the self-assembly (200 ps to 10 ns), indicating the collapse of the PAs. Ultimately, interactions among the hydrophilic amino acids and water enabled the formation of a stable fiber consisting of hydrophobic tails of the PA in the core and the hydrophilic groups in direct contact with water.¹¹²

3.3.2. Self-Assembly of FA32. As another application, we studied the effect of concentration on the final self-assembled structures of the FA32 peptide–oligopeptide amphiphiles composed of 32 amino acids, divided into three blocks with hydrophobic, amphiphilic, and hydrophilic properties—in water. Experimentally, this peptide is known to self-assemble into cationic micelles at low concentrations.^{113,114} The self-assembly of FA32 into micelles has been reported to be driven by the hydrophobic interactions between hydrophobic side chains in the amino acids.⁸² Our FF parameters allowed us to adequately model the FA32 peptide for comprehensive characterization of the PA self-assembly and the resulting equilibrium structure. The process of self-assembly of peptides at varying concentrations was examined by analyzing the simulation trajectory to calculate several parameters, including the number of clusters, aggregation number, as well as radii of the micelles. Furthermore, RDFs between water and hydrophobic, amphiphilic, and hydrophilic blocks of the molecules were obtained to gain insight into the solvation behavior of the system. Initially, to study the self-assembly behavior of FA32, we performed simulations at two different concentrations (12 and 60 FA32 molecules in a 180 Å box) using the nonbonded interaction parameters between amino acids and water obtained from the LB combining rule. Our simulations resulted in the formation of micelles in both cases. However, we observed that the hydrophobic portion of the molecules (Ala–Phe) resided at the surface of the micelles, while the hydrophilic portion (Lys) was located in the core of the micelle structure. This further suggested that the parameters obtained from the LB combining rule are not suitable to perform these self-assembly simulations. For a visual representation of the final structures obtained from the self-assembly simulations, please refer to Figure S8. This insightful

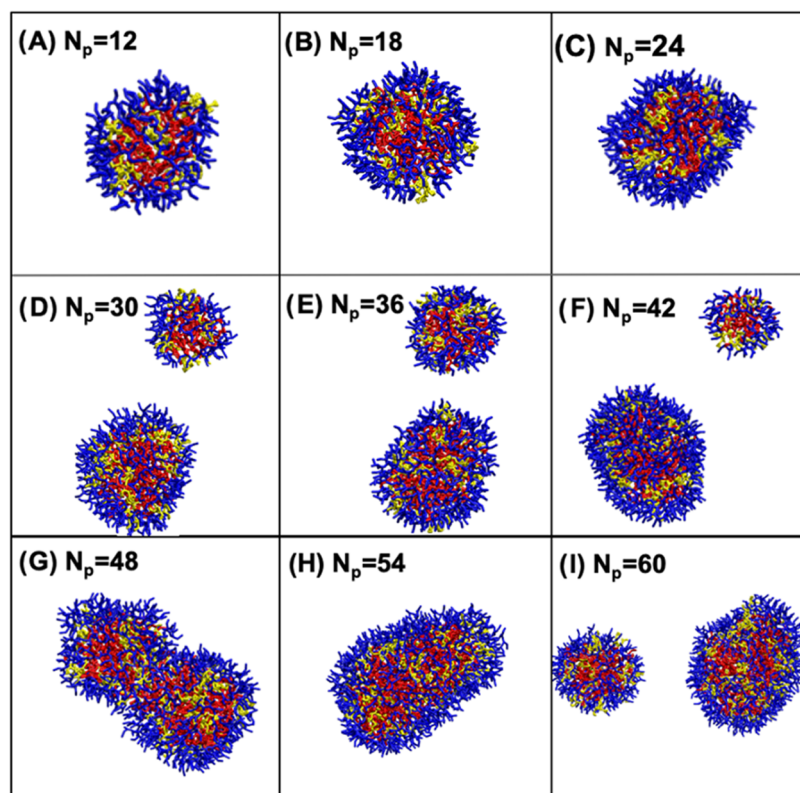


Figure 8. Final snapshots of the FA32 peptides at different concentrations containing 12, 18, 24, 30, 36, 42, 48, 54, and 60 molecules in a 180 Å water box. Ala and phe are represented by red beads. Amphiphilic his and hydrophilic lys are shown by yellow and blue beads, respectively. Water molecules are not shown for clarity.

result highlights the need of fine-tuning the nonbonded parameters between amino acids and water CG beads.

Subsequently, we conducted the self-assembly simulation using our newly developed parameters. After the initial few nanoseconds of self-assembly using homogeneously dispersed systems of FA32, neighboring peptides were found to aggregate into small clusters, driven by hydrophobic interactions. This trend continued as the simulation progressed, with the small clusters merging into larger ones that concentrated the hydrophobic residues at the core of these clusters. By the time the simulation reached 2–2.5 μ s, systems at different concentrations contained one or two micelles, formed due to the aggregation of multiple clusters, and this number remained consistent until the end of the simulation. Overall, for all of the concentrations, the assembly process was observed to involve three distinct stages: the amalgamation of small clusters, followed by the creation of large clusters, and ultimately the formation of micelles. Note, similar observations have been reported for the self-assembly studies performed using the MARTINI model.⁸² To facilitate a more intuitive understanding of the assembly process, visual aids were employed. Specifically, Figure 7 represents snapshots of 18 peptides in a 180 Å box at different time intervals. Snapshots of the final self-assembled structures at different concentrations are also shown in Figure 8. Movies of the assembly process for all of the systems at different concentrations are available in the SI. Movies are titled as “FA32-”, followed by the precise number of molecules utilized in the system across various concentrations (FA32-12.avi, FA32-18.avi, FA32-24.avi, FA32-30.avi, FA32-36.avi, FA32-42.avi, FA32-48.avi, FA32-54.avi, FA32-60.avi).

In order to study the evolution of the formation of clusters in the systems, we also conducted the DBSCAN analysis for all of the self-assembly simulations (see Section S3).^{106,107} As shown in Figure 9, we found that the number of clusters (N_c) decreased as the simulation time progressed, ultimately leading to the formation of one or two micelles. We observed a constant value of N_c for the last 1–1.5 μ s of CG MD simulation time, suggesting the stability of the self-assembled structures in the equilibrium state. Furthermore, we determined the radii of the assembled micelles by calculating their radii of gyration for the last 10,000 frames, corresponding to the last 100 ns of the simulations. We tabulated the number of self-assembled micelles, their aggregation number, and their radius in Table 3.

Generally, we observed that the radius of the micelles increased with an increase in the concentration. For instance, at a concentration of 18 molecules, the micelle radius was measured to be 22.93 ± 0.05 Å, while at a concentration of 48 molecules, the radius was 36.54 ± 0.13 Å. However, the radii of micelles at concentrations 48 and 54 molecules were higher than that of the concentration 60 molecules. This could be due to the formation of a single micelle at concentrations 48 and 54, whereas two micelles are observed at concentration 60. Note that similar to the computational study by Thota et al. using Martini FF,⁸² the radii of the micelles formed in our study are smaller compared to the experimentally reported values of ~ 50 nm (experimentally reported hydrodynamic particle size—the diameter of a hypothetical hard sphere that diffuses with the same speed as the particle being measured¹¹⁵—is ~ 102 nm).^{85,114}

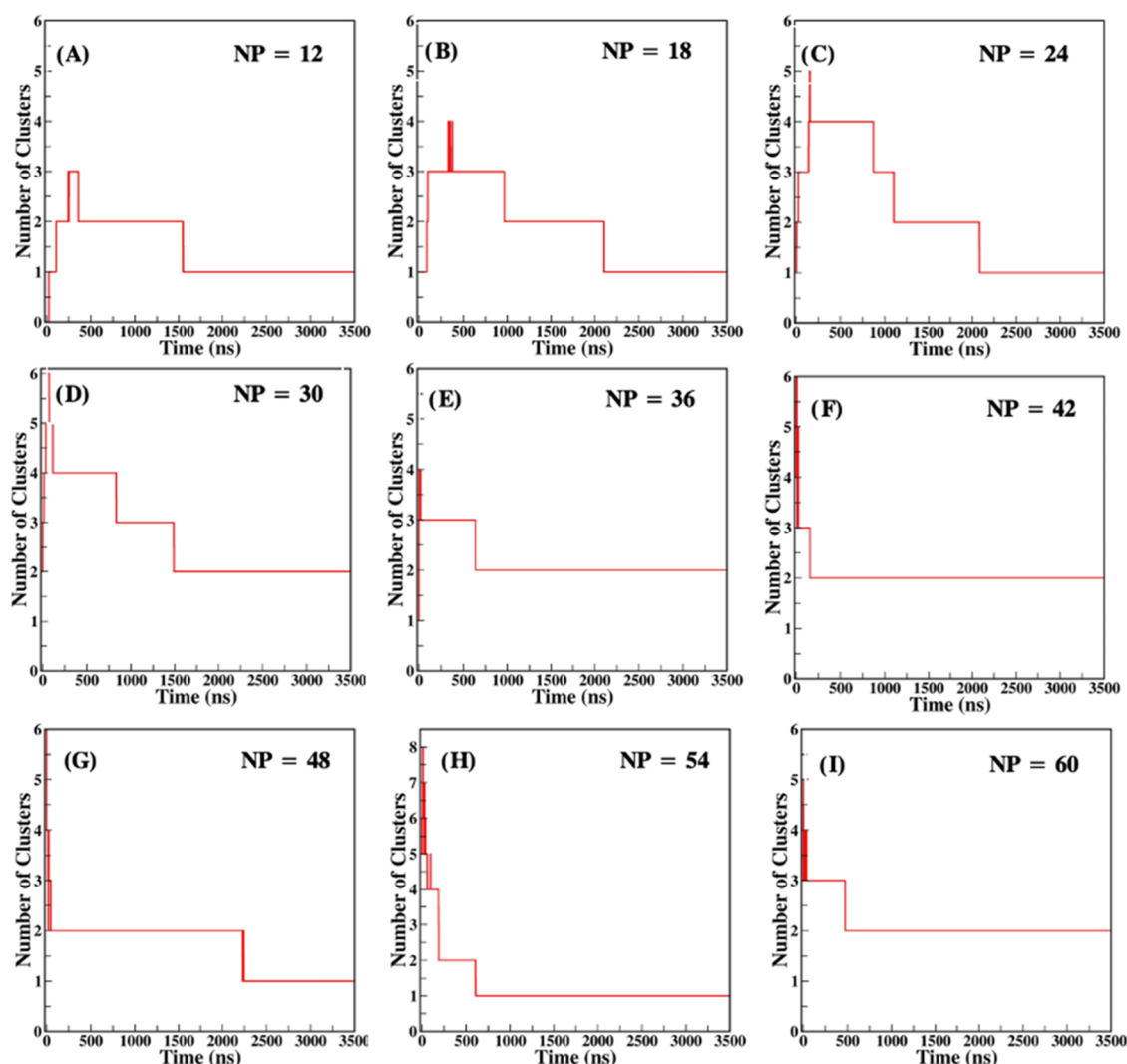


Figure 9. Number of clusters of polypeptide molecules versus time for the self-assembly process of FA32 at different concentrations, including 12, 18, 24, 36, 42, 48, 54, and 60 molecules.

Table 3. Number of Initial PAs, Number of Micelles, Radius, and Aggregation Number of Micelles in Nine Different Self-Assembly Simulations Containing Different Numbers of PA Molecules in the System

number of PAs	number of micelles	radius (Å)	aggregation number
12	1	20.30 ± 0.092	12
18	1	22.93 ± 0.05	18
24	1	25.34 ± 0.16	24
30	2	24.62 ± 0.07	22
		17.71 ± 0.09	8
36	2	25.31 ± 0.05	24
		20.14 ± 0.10	12
42	2	29.26 ± 0.062	36
		16.03 ± 0.09	6
48	1	36.54 ± 0.13	48
54	1	34.80 ± 0.11	54
60	2	30.88 ± 0.052	43
		22.15 ± 0.068	17

Similar to the radius, the aggregation number, representing the number of peptides within each micelle, increased with concentration. At a concentration of 12 molecules, the

aggregation number was 12, while at a concentration of 42 molecules, it increased to 36. The phenomenon of an increase in radius and aggregation number with increasing concentration has also been previously reported.⁸² As the concentration of amphiphilic polypeptides increased, we observed that the number of micelles generally increased. Specifically, at a concentration of 12 molecules, only one micelle was observed, while at higher concentrations of 30, 36, 42, and 60 molecules, the system formed two micelles. Interestingly, at concentrations 48 and 54 molecules, initially, two micelles were formed, but they aggregated together to form a single larger micelle. Thota et al. have also reported a similar complex relationship between the number of PA molecules and the number of assembled micelles.⁸²

Finally, to gain insight into the hydration behavior of self-assembled structures, we conducted RDF analysis. We used the last 1000 frames corresponding to the last 100 ns of the simulation to evaluate the hydrophilic behavior of the developed CG models. As shown in Figure 10, the RDF analysis for the system containing 18 polypeptides revealed that hydrophilic blocks of the polypeptides (lysine) are significantly more hydrated than amphiphilic (histidine) and hydrophobic blocks (alanine–phenylalanine). We observed a

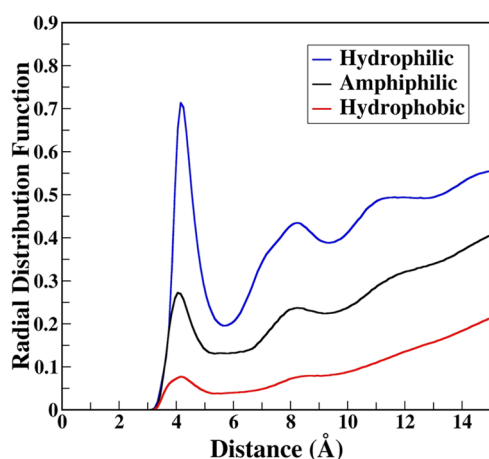


Figure 10. RDF plot calculated between hydrophilic amino acids (lys), amphiphilic amino acids (his), and hydrophobic residues (ala–phe) and water for a system containing 18 polypeptides.

similar trend for all other self-assembly systems, as shown in Figure S9. This result demonstrates the accuracy of capturing the hydrophilic behavior of the developed CG models and indicates that the hydrophilic amino acids were present at the surface of micelles, forming the micellar shell. Furthermore, the hydrophobic amino acids formed the core of the micelles, covered by the amphiphilic peptides, and exhibited the smallest peaks in the RDF profiles.

4. CONCLUSIONS

We developed and validated nonbonded interaction parameters between CG models of 20 amino acids and 1-site CG water model by reproducing the Gibbs hydration free energies of dipeptides obtained from all-atom simulations. These nonbonded interaction parameters were represented using the 12–6 LJ potential form where σ values were obtained using the LB combining rule, and ϵ values were developed using the ANN-assisted PSO method. The developed CG models reproduced the target Gibbs hydration free energies of dipeptides with errors less than 5%. The newly developed parameters were validated by calculating the hydration free energies of 18 analogues of amino acid side chains and could reproduce the experimental values for the majority of the analogues (14 out of 18) with errors less than 10% indicating their transferability to other small molecules.

Additionally, to investigate the applicability of our developed models, self-assembly simulations of two peptide amphiphiles, namely, c16-AHL3K3-CO₂H and FA32, were performed. These simulations resulted in experimentally observed fiber and micelle structures for c16-AHL3K3-CO₂H and FA32, respectively. The aggregation number and diameters of micelles and fibers were found to be in very good agreement with previous experimental and computational studies. In general, the visual and machine learning-based analysis suggested that the self-assembly process involved three distinct stages: the formation of small clusters, followed by their amalgamation into large clusters, and ultimately the formation of micelles. The RDF plots showed a strong structural correlation between water and the hydrophobic portion of the PAs at the early stages of self-assembly. With time, a reduction in RDF peaks between water and hydrophobic groups was observed, causing the hydrophobic collapse of PAs and their aggregation into micelles. Hydrophilic blocks of

polypeptides were found to be significantly more hydrated than amphiphilic and hydrophobic blocks forming the shell of micelles and fibers.

Overall, the optimized parameters developed in this work can be used to simulate various processes involving amino acids, paving the way for new insights into the mechanisms governing them.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biomac.3c00441>.

CG representation of amino acids (Table S1); CG representation of dipeptides (Table S2); particle swarm optimization (PSO) algorithm (Section S1); artificial neural network (ANN) (Section S2); flowchart of PSO algorithm (Figure S1); schematic of the ANN model (Figure S2); CG representation of amino acid analogues (Table S3); epsilon and $R_{\min}$ values of hydrocarbon CG beads (Table S4); Gibbs hydration free energies of AA and CG simulation of dipeptides (Figures S3 and S4, Table S5); bead density profiles of the hydrophobic tail of PA molecules (Figure S5); description of DBSCAN analysis (Section S3); RDF plot between the center of mass of the hydrophobic tail of FA32 molecules (Figure S6); ellipse fitting for diameter (Section S4); ellipse fitting algorithm for fiber diameter estimation (Figure S7); snapshots of FA32 peptides at two concentrations (Figure S8); RDF plots of self-assembly of polypeptides (Figure S9); and parameter file for conducting the CG MD simulation (PDF)

c16-AHL3K3-CO₂H (AVI)

FA32-12 (AVI)

FA32-18 (AVI)

FA32-24 (AVI)

FA32-30 (AVI)

FA32-36 (AVI)

FA32-42 (AVI)

FA32-48 (AVI)

FA32-54 (AVI)

FA32-60 (AVI)

■ AUTHOR INFORMATION

Corresponding Author

Sanket A. Deshmukh – Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; orcid.org/0000-0001-7573-0057; Email: sanketad@vt.edu

Authors

Esmat Mohammadi – Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; orcid.org/0009-0006-2631-4990

Soumil Y. Joshi – Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; orcid.org/0000-0003-1531-0098

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.biomac.3c00441>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-193352S. S.A.D. acknowledges the support from NSF CAREER Award (Award No. DMR-CMMT-2047743). The authors acknowledge Advanced Research Computing (ARC) at Virginia Tech for computational resources.

■ REFERENCES

- (1) Lopez, M. J.; Mohiuddin, S. S. *Biochemistry, Essential Amino Acids*. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2022.
- (2) Nelson, D. L.; Lehninger, A. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; Macmillan, 2008.
- (3) Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level*; John Wiley & Sons, 2016.
- (4) Ball, P. More than a Bystander. *Nature* **2011**, *478*, 467–468.
- (5) Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. Analytical Tools for Characterizing Biopharmaceuticals and the Implications for Biosimilars. *Nat. Rev. Drug Discovery* **2012**, *11*, 527–540.
- (6) Madeira, P. P.; Bessa, A.; Álvares-Ribeiro, L.; Aires-Barros, M. R.; Rodrigues, A. E.; Uversky, V. N.; Zaslavsky, B. Y. Amino Acid/water Interactions Study: A New Amino Acid Scale. *J. Biomol. Struct. Dyn.* **2014**, *32*, 959–968.
- (7) Ponting, C. P.; Russell, R. R. The Natural History of Protein Domains. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 45–71.
- (8) Davis, N. G.; Model, P. An Artificial Anchor Domain: Hydrophobicity Suffices to Stop Transfer. *Cell* **1985**, *41*, 607–614.
- (9) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. Sitting at the Edge: How Biomolecules Use Hydrophobicity to Tune Their Interactions and Function. *J. Phys. Chem. B* **2012**, *116*, 2498–2503.
- (10) Akdogan, Y.; Reichenwallner, J.; Hinderberger, D. Evidence for Water-Tuned Structural Differences in Proteins: An Approach Emphasizing Variations in Local Hydrophilicity. *PLoS One* **2012**, *7*, No. e45681.
- (11) Agre, P. Aquaporin Water Channels (Nobel Lecture). *Angew. Chem., Int. Ed.* **2004**, *43*, 4278–4290.
- (12) Deshmukh, S. A.; Solomon, L. A.; Kamath, G.; Fry, H. C.; Sankaranarayanan, S. K. R. S. Water Ordering Controls the Dynamic Equilibrium of Micelle-Fibre Formation in Self-Assembly of Peptide Amphiphiles. *Nat. Commun.* **2016**, *7*, No. 12367.
- (13) Xie, Y.; Li, N. K.; Singh, A.; Deshmukh, S. A.; Yingling, Y. G. A Comparison between the Lower Critical Solution Temperature Behavior of Polymers and Biomacromolecules. *PhysChem* **2022**, *2*, 52–71.
- (14) Wang, Y.; An, Y.; Shmidov, Y.; Bitton, R.; Deshmukh, S. A.; Matson, J. B. A Combined Experimental and Computational Approach Reveals How Aromatic Peptide Amphiphiles Self-Assemble to Form Ion-Conducting Nanohelices. *Mater. Chem. Front.* **2020**, *4*, 3022–3031.
- (15) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589*, 2477–2486.
- (16) Liao, H.-S.; Lin, J.; Liu, Y.; Huang, P.; Jin, A.; Chen, X. Self-Assembly Mechanisms of Nanofibers from Peptide Amphiphiles in Solution and on Substrate Surfaces. *Nanoscale* **2016**, *8*, 14814–14820.
- (17) Cote, Y.; Fu, I. W.; Dobson, E. T.; Goldberger, J. E.; Nguyen, H. D.; Shen, J. K. Mechanism of the pH-Controlled Self-Assembly of Nanofibers from Peptide Amphiphiles. *J. Phys. Chem. C* **2014**, *118*, 16272–16278.
- (18) Cui, H.; Webber, M. J.; Stupp, S. I. Self-Assembly of Peptide Amphiphiles: From Molecules to Nanostructures to Biomaterials. *Biopolymers* **2010**, *94*, 1–18.
- (19) Wei, G. *Self-Assembled Bio-Nanomaterials: Synthesis, Characterization, and Applications*; MDPI, 2020.
- (20) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. Self-Assembly and Mineralization of Peptide-Amphiphile Nanofibers. *Science* **2001**, *294*, 1684–1688.
- (21) Velichko, Y. S.; Stupp, S. I.; de la Cruz, M. O. Molecular Simulation Study of Peptide Amphiphile Self-Assembly. *J. Phys. Chem. B* **2008**, *112*, 2326–2334.
- (22) Zhao, X.; Pan, F.; Xu, H.; Yaseen, M.; Shan, H.; Hauser, C. A. E.; Zhang, S.; Lu, J. R. Molecular Self-Assembly and Applications of Designer Peptide Amphiphiles. *Chemical Society Reviews* **2010**, *39*, 3480.
- (23) Hendricks, M. P.; Sato, K.; Palmer, L. C.; Stupp, S. I. Supramolecular Assembly of Peptide Amphiphiles. *Acc. Chem. Res.* **2017**, *50*, 2440–2448.
- (24) An, Y.; Bejagam, K. K.; Deshmukh, S. A. Development of Transferable Nonbonded Interactions between Coarse-Grained Hydrocarbon and Water Models. *J. Phys. Chem. B* **2019**, *123*, 909–921.
- (25) Joshi, S. Y.; Singh, S.; Bejagam, K. K.; Deshmukh, S. A. Dehydration of Polymer Chains Initiates Graphene Folding in Water. *Carbon* **2021**, *180*, 244–253.
- (26) Joshi, S. Y.; Singh, S.; Deshmukh, S. A. Coarse-Grained Molecular Dynamics Integrated with Convolutional Neural Network for Comparing Shapes of Temperature Sensitive Bottlebrushes. *npj Comput. Mater.* **2022**, *8*, No. 45.
- (27) Dingreville, R.; Karnesky, R. A.; Puel, G.; Schmitt, J.-H. Review of the Synergies between Computational Modeling and Experimental Characterization of Materials across Length Scales. *J. Mater. Sci.* **2016**, *51*, 1178–1203.
- (28) Joshi, S. Y.; Deshmukh, S. A. A Review of Advancements in Coarse-Grained Molecular Dynamics Simulations. *Mol. Simul.* **2021**, *47*, 786–803.
- (29) Jiang, Z.; He, J.; Deshmukh, S. A.; Kanjanaboos, P.; Kamath, G.; Wang, Y.; Sankaranarayanan, S. K. R. S.; Wang, J.; Jaeger, H. M.; Lin, X.-M. Subnanometre Ligand-Shell Asymmetry Leads to Janus-like Nanoparticle Membranes. *Nat. Mater.* **2015**, *14*, 912–917.
- (30) Rakers, C.; Bermudez, M.; Keller, B. G.; Mortier, J.; Wolber, G. Computational Close up on Protein-Protein Interactions: How to Unravel the Invisible Using Molecular Dynamics Simulations? *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015**, *5*, 345–359.
- (31) Yuan, C.; Li, S.; Zou, Q.; Ren, Y.; Yan, X. Multiscale Simulations for Understanding the Evolution and Mechanism of Hierarchical Peptide Self-Assembly. *Phys. Chem. Chem. Phys.* **2017**, *19*, 23614–23631.
- (32) Negri, A.; Marco, E.; Damborsky, J.; Gago, F. Stepwise Dissection and Visualization of the Catalytic Mechanism of Haloalkane Dehalogenase LinB Using Molecular Dynamics Simulations and Computer Graphics. *J. Mol. Graphics Modell.* **2007**, *26*, 643–651.
- (33) Mohammadi, E.; Joshi, S. Y.; Deshmukh, S. A. A Review of Computational Studies of Bottlebrush Polymers. *Comput. Mater. Sci.* **2021**, *199*, No. 110720.
- (34) Wei, S.; Knotts, T. A., IV A Coarse Grain Model for Protein-Surface Interactions. *J. Chem. Phys.* **2013**, *139*, No. 095102.
- (35) Wu, J.; Zhen, X.; Shen, H.; Li, G.; Ren, P. Gay-Berne and Electrostatic Multipole Based Coarse-Grain Potential in Implicit Solvent. *J. Chem. Phys.* **2011**, *135*, No. 155104.
- (36) Heath, A. P.; Kavrakli, L. E.; Clementi, C. From Coarse-Grain to All-Atom: Toward Multiscale Analysis of Protein Landscapes. *Proteins* **2007**, *68*, 646–661.
- (37) Genheden, S. Solvation Free Energies and Partition Coefficients with the Coarse-Grained and Hybrid All-Atom/coarse-Grained MARTINI Models. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 867–876.
- (38) Bereau, T.; Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *J. Chem. Theory Comput.* **2015**, *11*, 2783–2791.
- (39) Singh, G.; Tieleman, D. P. Using the Wimley–White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields:

- (81) Fry, H. C.; Garcia, J. M.; Medina, M. J.; Ricoy, U. M.; Gosztola, D. J.; Nikiforov, M. P.; Palmer, L. C.; Stupp, S. I. Self-Assembly of Highly Ordered Peptide Amphiphile Metalloporphyrin Arrays. *J. Am. Chem. Soc.* **2012**, *134*, 14646–14649.
- (82) Thota, N.; Luo, Z.; Hu, Z.; Jiang, J. Self-Assembly of Amphiphilic Peptide (AF)6H5K15: Coarse-Grained Molecular Dynamics Simulation. *J. Phys. Chem. B* **2013**, *117*, 9690–9698.
- (83) Thota, N.; Jiang, J. Self-Assembly of Amphiphilic Peptide (AF)6H5K15 Derivatives: Roles of Hydrophilic and Hydrophobic Residues. *J. Phys. Chem. B* **2014**, *118*, 2683–2692.
- (84) Fry, H. C.; Liu, Y.; Dimitrijevic, N. M.; Rajh, T. Photoinitiated [corrected] Charge Separation in a Hybrid Titanium Dioxide Metalloporphyrin Peptide Material. *Nat. Commun.* **2014**, *5*, No. 4606.
- (85) Wiradharma, N.; Tong, Y. W.; Yang, Y.-Y. Design and Evaluation of Peptide Amphiphiles with Different Hydrophobic Blocks for Simultaneous Delivery of Drugs and Genes. *Macromol. Rapid Commun.* **2010**, *31*, 1212–1217.
- (86) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.
- (87) Waldman, M.; Hagler, A. T. New Combining Rules for Rare Gas van Der Waals Parameters. *J. Comput. Chem.* **1993**, *14*, 1077–1084.
- (88) Song, W.; Rossky, P. J.; Maroncelli, M. Modeling Alkane Perfluoroalkane Interactions Using All-Atom Potentials: Failure of the Usual Combining Rules. *J. Chem. Phys.* **2003**, *119*, 9145–9162.
- (89) Yesudasan, S.; Averett, R.; Chacko, S. Machine Learned Coarse Grain Water Models for Evaporation Studies *Preprints* 2020, DOI: 10.20944/preprints202007.0126.v1.
- (90) Das, S.; Baghel, V. S.; Roy, S.; Kumar, R. A Molecular Dynamics Study of Model SI Clathrate Hydrates: The Effect of Guest Size and Guest-Water Interaction on Decomposition Kinetics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 9509–9518.
- (91) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849–855.
- (92) In, Y.; Chai, H. H.; No, K. T. A Partition Coefficient Calculation Method with the SFED Model. *J. Chem. Inf. Model.* **2005**, *45*, 254–263.
- (93) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23*, 517–529.
- (94) Plyasunov, A. V.; Shock, E. L. Thermodynamic Functions of Hydration of Hydrocarbons at 298.15 K and 0.1 MPa. *Geochim. Cosmochim. Acta* **2000**, *64*, 439–468.
- (95) Jorgensen, W. L.; McDonald, N. A. Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyridine and Diazenes. *J. Mol. Struct.: THEOCHEM* **1998**, *424*, 145–155.
- (96) McDonald, N. A.; Jorgensen, W. L. Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyrrole, Furan, Diazoles, and Oxazoles. *J. Phys. Chem. B* **1998**, *102*, 8049–8059.
- (97) Conway, O.; An, Y.; Bejagam, K. K.; Deshmukh, S. A. Development of Transferable Coarse-Grained Models of Amino Acids. *Mol. Syst. Des. Eng.* **2020**, *5*, 675–685.
- (98) Lee, O.-S.; Cho, V.; Schatz, G. C. Modeling the Self-Assembly of Peptide Amphiphiles into Fibers Using Coarse-Grained Molecular Dynamics. *Nano Lett.* **2012**, *12*, 4907–4913.
- (99) Fu, I. W.; Markegard, C. B.; Chu, B. K.; Nguyen, H. D. The Role of Electrostatics and Temperature on Morphological Transitions of Hydrogel Nanostructures Self-Assembled by Peptide Amphiphiles Via Molecular Dynamics Simulations. *Adv. Healthcare Mater.* **2013**, *2*, 1388–1400.
- (100) Lee, O.-S.; Stupp, S. I.; Schatz, G. C. Atomistic Molecular Dynamics Simulations of Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Am. Chem. Soc.* **2011**, *133*, 3677–3683.
- (101) Shimada, T.; Lee, S.; Bates, F. S.; Hotta, A.; Tirrell, M. Wormlike Micelle Formation in Peptide-Lipid Conjugates Driven by Secondary Structure Transformation of the Headgroups. *J. Phys. Chem. B* **2009**, *113*, 13711–13714.
- (102) Yu, T.; Schatz, G. C. Free-Energy Landscape for Peptide Amphiphile Self-Assembly: Stepwise versus Continuous Assembly Mechanisms. *J. Phys. Chem. B* **2013**, *117*, 14059–14064.
- (103) Zaldivar, G.; Vemulapalli, S.; Udumula, V.; Conda-Sheridan, M.; Tagliazucchi, M. Self-Assembled Nanostructures of Peptide Amphiphiles: Charge Regulation by Size Regulation. *J. Phys. Chem. C* **2019**, *123*, 17606–17615.
- (104) Trent, A.; Marullo, R.; Lin, B.; Black, M.; Tirrell, M. Structural Properties of Soluble Peptide Amphiphile Micelles. *Soft Matter* **2011**, *7*, No. 9572.
- (105) Nagarajan, R.; Ruckenstein, E. Aggregation of Amphiphiles as Micelles or Vesicles in Aqueous Media. *J. Colloid Interface Sci.* **1979**, *71*, 580–604.
- (106) Ur Rehman, S.; Asghar, S.; Fong, S.; Sarasvady, S. DBSCAN: Past, Present, and Future, The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014); IEEE, 2014.
- (107) Schubert, E.; Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X. DBSCAN Revisited, Revisited. *ACM Trans. Database Syst.* **2017**, *42*, 1–21.
- (108) Wade, R. J.; Burdick, J. A. Advances in Nanofibrous Scaffolds for Biomedical Applications: From Electrospinning to Self-Assembly. *Nano Today* **2014**, *9*, 722–742.
- (109) Morén, A. K.; Khan, A. Surfactant Hydrophobic Effect on the Phase Behavior of Oppositely Charged Protein and Surfactant Mixtures: Lysozyme and Sodium Alkyl Sulfates. *Langmuir* **1998**, *14*, 6818–6826.
- (110) Kronberg, B.; Costas, M.; Silveston, R. Thermodynamics of the Hydrophobic Effect in Surfactant Solutions: Micellization and Adsorption. *Pure Appl. Chem.* **1995**, *67*, 897–902.
- (111) Spolar, R. S.; Ha, J. H.; Record, M. T. Hydrophobic Effect in Protein Folding and Other Noncovalent Processes Involving Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8382–8385.
- (112) Yu, T.; Schatz, G. C. Free Energy Profile and Mechanism of Self-Assembly of Peptide Amphiphiles Based on a Collective Assembly Coordinate. *J. Phys. Chem. B* **2013**, *117*, 9004–9013.
- (113) Chen, S.; Cheng, S.-X.; Zhuo, R.-X. Self-Assembly Strategy for the Preparation of Polymer-Based Nanoparticles for Drug and Gene Delivery. *Macromol. Biosci.* **2011**, *11*, 576–589.
- (114) Wiradharma, N.; Tong, Y. W.; Yang, Y.-Y. Self-Assembled Oligopeptide Nanostructures for Co-Delivery of Drug and Gene with Synergistic Therapeutic Effect. *Biomaterials* **2009**, *30*, 3100–3109.
- (115) Batchelor, G. K. Brownian Diffusion of Particles with Hydrodynamic Interaction. *J. Fluid Mech.* **1976**, *74*, 1–29.