

Coarse-Grained Molecular Simulation of the Hierarchical Self-Assembly of π -Conjugated Optoelectronic Peptides

Rachael A. Mansbach[†] and Andrew L. Ferguson^{*,‡,§,ID}

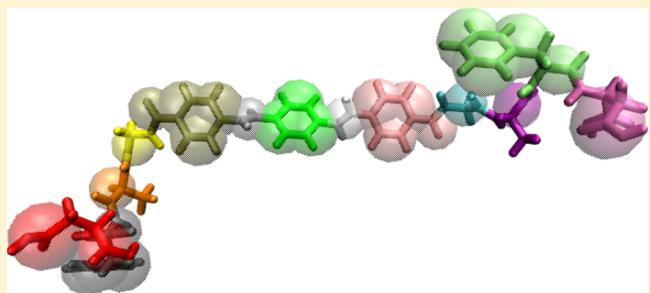
[†]Department of Physics, University of Illinois at Urbana-Champaign, 1110 W Green Street, Urbana, Illinois 61801, United States

[‡]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, 1304 W Green Street, Urbana, Illinois 61801, United States

[§]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Self-assembled aggregates of peptides containing aromatic groups possess optoelectronic properties that make them attractive targets for the fabrication of biocompatible electronics. Molecular-level understanding of the influence of microscopic peptide chemistry on the properties of the aggregates is vital for rational peptide design. In this study, we construct a coarse-grained model of Asp-Phe-Ala-Gly-OPV3-Gly-Ala-Phe-Asp (DFAG-OPV3-GAFD) peptides containing OPV3 (distyrylbenzene) π -conjugated cores explicitly parameterized against all-atom calculations and perform molecular dynamics simulations of the self-assembly of hundreds of molecules over hundreds of nanoseconds. We observe a hierarchical assembly mechanism, wherein approximately two to eight peptides assemble into stacks with aligned aromatic cores that subsequently form elliptical aggregates and ultimately a branched network with a fractal dimensionality of ~ 1.5 . The assembly dynamics are well described by a Smoluchowski coagulation process, for which we extract rate constants from the molecular simulations to both furnish insight into the microscopic assembly kinetics and extrapolate our aggregation predictions to time and length scales beyond the reach of molecular simulation. This study presents new molecular-level understanding of the morphology and dynamics of the spontaneous self-assembly of DFAG-OPV3-GAFD peptides and establishes a systematic protocol to develop coarse-grained models of optoelectronic peptides for the exploration and design of π -conjugated peptides with tunable optoelectronic properties.



1. INTRODUCTION

As the understanding of biological molecules and techniques to synthesize new materials continues to improve, there is growing interest in “designer” nanomaterials: organic materials that can be tuned at a microscopic level to produce desirable macroscopic properties. Proteins and polymers that display triggerable self-assembly into supramolecular aggregates have proven to be valuable components in the fabrication of such nanomaterials,¹ with a wide range of potential applications, including drug delivery, tissue regeneration,^{2–4} and as biominerization scaffolds for tooth and bone growth.^{5,6} Furthermore, “smart” self-assembling peptide-based materials have found multifarious applications in biosensing because of their ability to respond to diverse stimuli, including pH, temperature, metal ion concentration, enzymatic activity, and light bombardment.⁷ In other cases, such external stimuli have been used to tune the properties of the resultant aggregates for specific applications in materials design,⁸ biomedical diagnostics,⁹ and bioelectronics.¹⁰

Self-assembling π -conjugated peptides are easily synthesizable, water soluble, and biocompatible. Their optoelectronic functionality, due to the overlap and delocalization of π orbitals

within the supramolecular aggregates,¹¹ makes them well suited for bioelectronic applications, such as organic field-effect transistors, light-emitting diodes, and photovoltaic cells.^{12–16} They also display promising tunability, with the assembly behavior and emergent photophysical and electronic properties of the aggregates controllable by manipulation of the chemistry of the conjugated cores and side-chain moieties and the prevailing environmental conditions.¹³ Several experimental and computational studies have been performed on π -conjugated peptide systems containing oligophenylenevinylene (OPV) cores. It has been demonstrated that these systems can undergo β -sheet-like assembly into one-dimensional nanostructures, the optoelectronic properties of which can be tuned from excimeric-like to excitonic-like by changing the chemistry and symmetry of the molecules.^{16–19} In this article, we focus on the symmetric Asp-Phe-Ala-Gly-OPV3-Gly-Ala-Phe-Asp (DFAG-OPV3-GAFD) oligopeptide containing an OPV3 (distyrylbenzene) π -conjugated core, in which the N-to-C directionality of

Received: October 7, 2016

Revised: December 13, 2016

Published: January 24, 2017

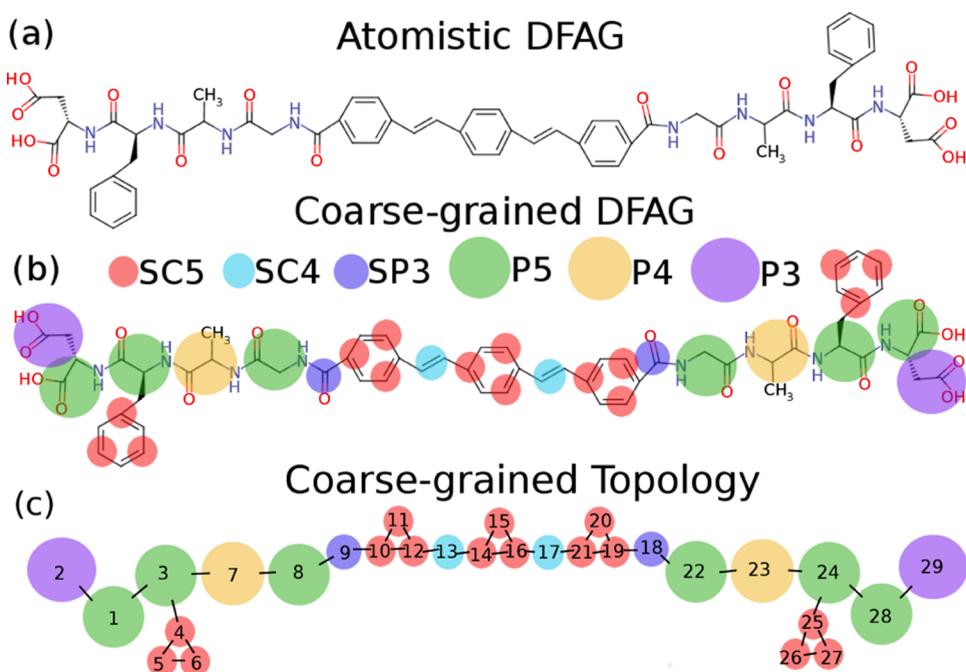


Figure 1. Atomistic and coarse-grained representations of the Asp-Phe-Ala-Gly-OPV3-Gly-Ala-Phe-Asp (DFAG-OPV3-GAFD) π -conjugated oligopeptide. These molecules are symmetric in the sense that each of the peptidic wings possess N-to-C directionality proceeding away from the π -conjugated core such that there are two C-termini.^{17,18} (a) Chemical structure of DFAG-OPV3-GAFD, produced using Marvin 15.12.7.0 (ChemAxon, 2015) (<http://www.chemaxon.com>). Under high pH conditions (pH $\gtrsim 10$), the distal aspartic acid residues are doubly negatively charged, prohibiting large-scale assembly due to electrostatic repulsion; under low pH conditions (pH $\lesssim 2$), protonation of the terminal residues leads to large-scale aggregation of the electrically neutral molecules.¹⁸ (b) Mapping of the DFAG-OPV3-GAFD oligopeptide from an all-atom representation to a coarse-grained representation under the Martini model.^{22,23} Colored circles represent different coarse-grained beads formed by lumping of atoms. Detailed descriptions of the Martini beads may be found in Monticelli et al.²² (c) Coarse-grained topology of the DFAG-OPV3-GAFD oligopeptide with bonds and bead numbering scheme implemented in the coarse-grained adaptation of the Martini force field that we explicitly reparameterize against all-atom calculations. Full details of the force field are provided in the Supporting Information.

the peptide wings proceed away from the OPV3 core such that the molecule possesses two C-termini (Figure 1a).^{17,18} Under high pH conditions, these peptides exist as dispersed monomers or small oligomers, with large-scale assembly prohibited by the electrostatic repulsion between the doubly negatively charged aspartic acid termini.²⁰ Under low pH conditions, protonation of the termini eliminates the electrostatic repulsion and triggers the assembly into long β -sheet-like aggregates mediated by hydrogen bonding, π - π stacking, and dispersion interactions between the peptide cores and side chains,^{15–17,19,20} producing structures similar to those of amyloid fibrils.²¹

Previous theoretical and computational work on the DFAG-OPV3-GAFD assembly has included simulations of dimer aggregation and dimerization potentials of mean force (PMF),¹⁶ simulation and characterization of the morphology of preformed one-dimensional nanostructures of tens of monomers,¹⁸ and the characterization of the early stages of assembly up to tens of nanoseconds using implicit solvent molecular dynamics simulations and a Markov state model.²⁰ These works shed valuable light on both the kinetics of early-stage assembly and the effects of chemistry on the morphology of preformed aggregates. No study to date has directly simulated the assembly of aggregates containing more than ~ 10 oligopeptides, and it is of interest to probe longer time and length scales to determine the morphology and dynamics of the assembly beyond these small-scale and early-stage events.

In this work, we construct a coarse-grained molecular model explicitly parameterized against all-atom calculations that

enables us to perform molecular dynamics simulations of the assembly of hundreds of monomers over hundreds of nanoseconds. The structure of this article is as follows. In Section 2, we describe the development of a coarse-grained DFAG-OPV3-GAFD model from all-atom calculations and the use of this model to perform coarse-grained molecular dynamics simulations. In Section 3, we report the results of our coarse-grained simulations, characterize the morphology of the self-assembly pathway, and extract microscopic rate constants to parameterize a Smoluchowski coagulation model of nonequilibrium dynamics of peptide assembly. In Section 4, we present our conclusions and outlook for future work.

2. THEORETICAL METHODS

In this section, we report the development of a coarse-grained model explicitly parameterized against all-atom calculations, describe the simulation of hundreds of coarse-grained monomers over hundreds of nanoseconds at different concentrations, and define the structural measures used to characterize cluster morphology.

2.1. Development of a Coarse-Grained Model through Atomistic and Coarse-Grained Molecular Dynamics Simulations. Molecular dynamics simulations provide a means to link the chemical details of self-assembling peptides to the emergent structural and functional properties of self-assembled aggregates, providing a route to molecular-level mechanistic understanding and rational materials design.¹ All-atom simulations of hundreds of thousands of atoms over microsecond time scales are computationally challenging.

Coarse-grained or multiphysics models reduce the number of degrees of freedom, providing the efficiency gains necessary to access the time and length scales required to robustly observe the assembly.^{24–26} Coarse-grained molecular dynamics simulations in which groups of atoms are lumped together into single beads have been employed to elucidate the structural and functional properties of large systems, including proteins,^{27,28} DNA,²⁹ lipids and lipid membranes,²⁴ and even bacterial flagella.³⁰

In this work, we develop a coarse-grained potential on the basis of the Martini potential,^{22,23} which is a popular coarse-grained model³¹ of proteins,^{23,32} lipids,³³ and carbohydrates³⁴ that has been profitably employed to study peptide self-assembly.^{35,36} The Martini force field is computationally efficient, transferable, and explicitly distinguishes between the 20 natural amino acid side chains,^{22,37} providing a good balance between computational efficiency and chemical realism that permits us to explicitly link oligopeptide sequence to the emergent assembly behavior. As described below, we construct coarse-grained Martini models for the DFAG-OPV3-GAFD oligopeptide and then recalibrate the Martini parameters against all-atom simulation data to develop a coarse-grained force field offering $\sim 25\times$ speedups over all-atom calculations. The approach we present offers a systematic means to develop efficient and accurate bespoke coarse-grained potentials, explicitly incorporating atomistic information.

2.1.1. All-Atom Molecular Dynamics Simulations for Coarse-Grained Model Parameterization. We employed the Gromacs 4.6 simulation suite³⁸ to perform all-atom molecular dynamics simulations of a single DFAG-OPV3-GAFD molecule solvated in water. Initial peptide configurations were constructed using the GlycoBioChem PRODRG2 server.³⁹ The peptides were prepared with protonated Asp termini corresponding to the electrically neutral, low-pH form of these molecules that have been experimentally demonstrated to exhibit a triggered self-assembly under acidic conditions.¹⁷ We modeled water explicitly using the TIP3P model⁴⁰ and treated the peptides using the CHARMM27 force field,⁴¹ supplemented with bonded parameters required to model the OPV3 core as described in Thurston et al.²⁰ Simulations were conducted in the NPT ensemble at 298 K and 1 bar using a Nosé–Hoover thermostat⁴² and Parrinello–Rahman barostat.⁴³ Three-dimensional periodic boundary conditions were employed. Initial atom velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K, and high-energy overlaps in initial configurations were eliminated by performing the steepest descent energy minimization to remove forces exceeding 500 kJ/mol nm. The equations of motion were numerically integrated using a leap-frog algorithm⁴⁴ with a 2 fs time step, and bond lengths were fixed using the LINCS algorithm to improve the efficiency.⁴⁵ Electrostatic interactions were treated using particle mesh Ewald with a real-space cutoff of 1.2 nm and a 0.12 nm Fourier grid spacing.⁴⁶ The Lennard–Jones interactions were shifted smoothly to zero at 1.2 nm, and Lorentz–Berthelot combining rules were used to determine the interaction parameters between unlike atoms.⁴⁷ Calculations were performed on 20 \times 2.2 GHz Intel Xeon E5-2430 CPU cores, achieving execution speeds of ~ 30 ns/day.

2.1.1.1. Isolated Molecules. A single DFAG-OPV3-GAFD molecule was placed in a $7 \times 7 \times 7$ nm³ box, which was sufficiently large to preclude direct interactions of the peptide with its own periodic image in the fully extended state. The peptide was then solvated to a density of 0.988 g/cm³ with

TIP3P water,⁴⁰ and an equilibration run of 10 ps was performed, after which time the density, energy, temperature, and pressure attained stable values. A production run of 340 ns was conducted to comprehensively sample the intramolecular degrees of freedom.

2.1.1.2. Monomer PMF. The end-to-end PMFs quantifying the free energy of a monomer as a function of the distance between its terminal carbon atoms were computed using umbrella sampling in the end-to-end distance.⁴⁸ The oligopeptide termini were restrained at intermolecular separations between 0.0 and 4.0 nm at intervals of 0.1–0.2 nm to assure a good overlap of neighboring umbrella window histograms. A harmonic restraining potential of 1000 kJ/mol nm² was employed in each window, and the system allowed to relax for 100 ps before a 10 ns production run. End-to-end distances were recorded and the PMF computed by self-consistent solution of the weighted histogram analysis method (WHAM) equations⁴⁹ to a tolerance of 10^{-6} in the estimated probability distribution, and uncertainties estimated by block averaging.

2.1.1.3. Dimerization PMF. Dimerization PMFs quantifying the free energy as a function of the center-of-mass separation between a pair of monomers were computed in umbrella sampling calculations, in which the monomer pair was restrained at separations between 0.0 and 5.0 nm at intervals of 0.1–0.2 nm using a harmonic potential with a force constant of 1000 kJ/mol nm². The largest separation considered was sufficient to observe a plateau in the PMF corresponding to a regime in which the monomers are effectively noninteracting. The dimerization PMF was computed by solving the WHAM equations⁴⁹ to a tolerance of 10^{-6} in the estimated probability distribution, and uncertainties were estimated by block averaging.

2.1.2. Coarse-Grained Mapping and Force-Field Parameterization. Coarse-grained models of the DFAG-OPV3-GAFD molecule were prepared in the low-pH state (pH $\lesssim 1$), in which the aspartic acid residues are fully protonated and the molecule is electrically neutral, and the high-pH state (pH $\gtrsim 10$), in which the aspartic acids are fully deprotonated and the molecule carries a net charge of -4 .^{18,20} We first describe the construction of the protonated (low-pH) coarse-grained model and then the modifications made for the deprotonated (high-pH) model. The construction of the coarse-grained model from all-atom simulation data proceeds in three stages: (i) mapping from all-atom to coarse-grained topology, (ii) definition of nonbonded (intermolecular) interactions between coarse-grained beads, and (iii) definition of bonded (intramolecular) interactions between the beads.

2.1.2.1. Coarse-Grained Mapping. A coarse-grained representation of the electrically neutral (low-pH) DFAG-OPV3-GAFD was constructed from the all-atom structure using the Martini model, in which approximately four heavy atoms are mapped to a single bead (Figure 1b).^{22,23} An initial molecular topology was built using the martiniize.py script available from the Martini website (<http://www.cgmartini.nl>),²³ which was subsequently modified to treat the aromatic rings as described in the Martini force field⁵⁰ and the peptide–core linker regions by following the guidelines on the Martini website (<http://www.cgmartini.nl/index.php/faq/topologies>) (Figure 1c).

2.1.2.2. Nonbonded Interactions. The nonbonded interactions were treated using the Martini force field (version 2.2)²³ and the nonpolarizable Martini water model.⁵¹

2.1.2.3. Bonded Interactions. Bonded parameters for the coarse-grained bond-stretching, angle-bending, and proper and

improper dihedrals were extracted from our 340 ns all-atom simulation trajectory using Boltzmann inversion to infer an effective coarse-grained potential from all-atom intramolecular distributions.^{26,52} By explicitly calculating bonded potentials in this manner, we developed a bespoke coarse-grained model that exhibited improved agreement with the all-atom calculations compared to the standard Martini parameters (see Supporting Information). A deficiency of coarse-grained models is that they frequently do not adequately reproduce the intramolecular secondary structure of the all-atom model (cf. the monomer collapse PMF of the unmodified Martini model in Figure 2a). In addition to reparameterizing each bonded Martini interaction, we follow Seo et al.³⁶ and apply proper dihedrals to every group of four consecutive beads in the backbone to enforce the backbone conformations observed in the all-atom calculations. The specific beads participating in these additional proper dihedrals and the parameters of the dihedral potential (cf. eq 1) are listed in Tables S5–S8.

To compute the effective coarse-grained potentials, we first located the hypothetical coarse-grained beads within the all-atom simulation trajectory as the centers of mass of the atoms mapped to each of the corresponding coarse-grained beads. We then compiled the histograms of the distribution functions for each coarse-grained bonded interaction over the course of the all-atom trajectory under this coarse-grained mapping and extracted the tabulated effective coarse-grained bonded potentials from these histograms. Both procedures were performed by using the direct Boltzmann inversion module in the Coarse-graining Toolkit of the Versatile Object-oriented Toolkit for Coarse-graining Applications software package (VOTCA-CSG).⁵³ On implementing these effective potentials in Gromacs, the bond-stretching and angle-bending potentials were represented directly as tabulated potentials produced by VOTCA, whereas the periodic proper dihedrals were represented in the form

$$V_d(\phi) = \sum_{n=1}^4 C_n [1 + \cos(n\phi)] + \sum_{n=0}^4 S_n \left[1 + \cos\left(n\phi - \frac{\pi}{2}\right) \right] \quad (1)$$

by constructing least-square fits in $\{C_n\}_{n=1}^4$ and $\{S_n\}_{n=0}^4$ to the tabulated potential using Matlab.⁵⁴ For convenience of representation in Gromacs, the improper dihedrals were represented using the same functional form as the proper dihedrals. The distribution functions for bond lengths within the rigid aromatic rings were observed to be extremely narrow, so—similar to the rigid coarse-grained model of benzene proposed by Marrink et al.⁵¹—we elected to treat aromatic bond lengths as hard constraints enforced by the LINCS algorithm.⁴⁵ The symmetric DFAG-OPV3-GAFD molecule contains several chemically identical bonded interactions, and we improved the sampling in our inversion procedure by averaging the calculated effective potentials over all equivalent bonded interactions.

2.1.2.4. Protonated (Low-pH) and Deprotonated (High-pH) Models. The low-pH state with fully protonated aspartic acid residues requires two modifications to construct a high-pH model containing deprotonated aspartic acid residues. First, the aspartic acid beads were converted from P3 and P5 semi-attractive beads to negatively charged Qa beads, corresponding to the Martini treatment of the change in the Asp protonation

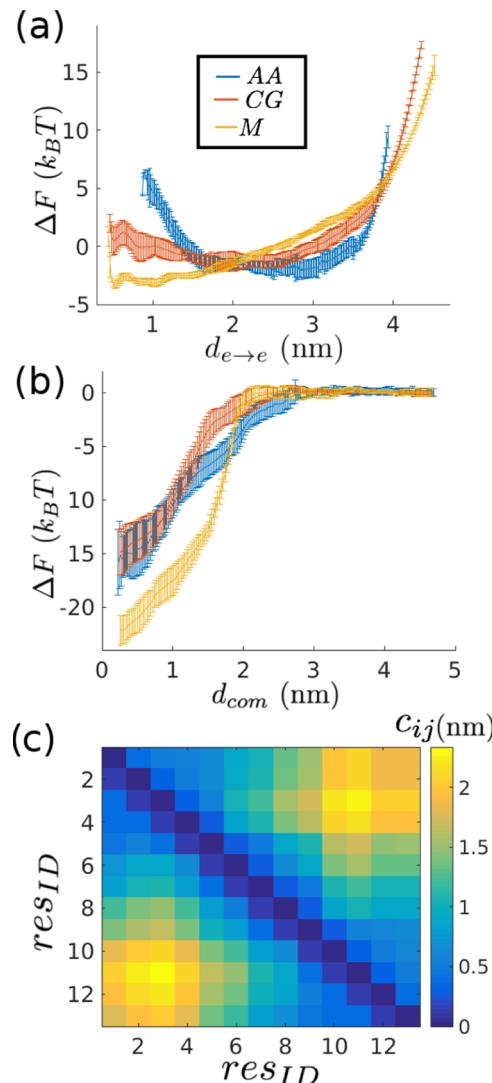


Figure 2. Validation of the reparameterized coarse-grained model. (a) Monomer collapse PMFs for all-atom (AA), reparameterized coarse-grained (CG), and original Martini (M) models. $d_{e \rightarrow e}$ represents the distance between the terminal carbon atoms or the beads containing the terminal carbon atoms and ΔF is the (relative) Gibbs free energy determined by umbrella sampling. The arbitrary additive constant specifying the vertical offset of each PMF profile was specified for the all-atom model such that ΔF averaged over the profile is zero and for the two coarse-grained models to minimize the pointwise root-mean-square distance (m_{rms}) relative to the all-atom profile (cf. eq 2). (b) Dimerization PMFs for the all-atom, reparameterized coarse-grained, and original Martini models. d_{com} represents the center-of-mass distance between the two monomers and ΔF is the (relative) Gibbs free energy determined by umbrella sampling. The arbitrary additive constant specifying the vertical offsets of the PMF profiles was specified by aligning the plateau regions at large separations corresponding to effectively noninteracting monomers to $\Delta F = 0 k_B T$. (c) Time-averaged contact maps of the all-atom (lower triangle) and reparameterized coarse-grained (upper triangle) models.

state.²² Second, water was treated by the Martini polarizable water model, which, despite its higher computational cost, is a more realistic model for systems containing charged particles, where polarization effects are anticipated to be important.⁵⁵

2.1.3. Coarse-Grained Molecular Dynamics Simulations for Model Parameterization. We now describe the coarse-

grained molecular dynamics simulations conducted to test the performance of the coarse-grained model.

2.1.3.1. Isolated Molecules. To assess the quality of the original Martini and reparameterized coarse-grained models, we performed coarse-grained molecular dynamics simulations of a single monomer in solution. Calculations were conducted in the Gromacs 4.6 simulation suite.⁵⁶ A single DFAG-OPV3-GAFD peptide was placed in a $7 \times 7 \times 7$ nm³ box, which was sufficiently large to preclude direct interactions of the peptide with its own periodic image even in its fully extended state. The peptide was then solvated to a density of 0.988 g/cm³ with Martini nonpolarizable⁵¹ or Martini polarizable⁵⁵ water, depending on which model was being tested. Simulations were conducted in the NPT ensemble at 298 K and 1 bar, employing a velocity rescaling thermostat,⁵⁷ a Parrinello–Rahman barostat,⁴³ and three-dimensional periodic boundary conditions. Initial velocities were randomly drawn from a Maxwell–Boltzmann distribution at 298 K, and high-energy overlaps in the initial configurations were eliminated by performing the steepest descent energy minimization to remove forces exceeding 1000 kJ/mol nm. Equations of motion were numerically integrated using a leap-frog algorithm⁴⁴ with a model-dependent time step, and constrained bond lengths were fixed using the LINCS algorithm.⁴⁵ The Lennard-Jones interactions were shifted smoothly to zero at 1.1 nm. Electrostatic interactions were treated using reaction-field electrostatics and a relative electrostatic screening of 2.5 for polarizable water and 15 for nonpolarizable water. An equilibration run of 10 ps was performed, after which time, the density, energy, temperature, and pressure attained stable values. Following equilibration, a production run of 340 ns was performed to comprehensively sample the intramolecular degrees of freedom. An average calculation was performed on 6 × 2.2 GHz Intel Xeon E5-2430 CPU cores, achieving execution speeds of ~1100 ns/day.

2.1.3.2. Monomer and Dimerization PMF. PMFs for the reparameterized coarse-grained model and the original Martini model were computed in an identical manner to those described for the all-atom model in Section 2.1.1, with umbrella potentials applied to the beads containing the terminal carbon atoms (Figure 1).

2.1.4. Validation of Coarse-Grained Model. We assess and validate the performance of the reparameterized coarse-grained model in reproducing the bonded distribution functions, all-atom monomer collapse PMF, dimerization PMF, and contact map. A complete list of the coarse-grained bonded and nonbonded parameters is provided in Tables S1–S8, the tabulated force-field files b0–b8.txt and a0–a12.txt, and the DFAG coarse-grained topology topology.txt in Supporting Information.

2.1.4.1. Bonded Distributions. It is important that the bonded distributions of a coarse-grained model be in good agreement with the all-atom results to accurately reproduce the internal bonded structure of the molecule. Coarse-grained simulations of the original Martini model showed a relatively poor agreement with the all-atom calculations, and we improved the quality of the coarse-grained model by reparameterizing the bonded interactions using the Boltzmann inversion procedure detailed in Section 2.1.2. In Figures S1–S5, we present a comparison between the 33 bond, 24 angle, 15 proper dihedral, and 2 improper dihedral distributions sampled by the all-atom and reparameterized coarse-grained models. In all cases, the reparameterized coarse-grained model adequately

recapitulates the range and character of the bonded distribution functions, with the exception of 6/33 bonds (Figure S1e–j) and 9/24 angles (Figure S3a,d,e,l,m,q–t), where it fails to precisely reproduce bimodalities observed in the all-atom calculations. This should be regarded as a deficiency of the reparameterized coarse-grained model that could be ameliorated by more sophisticated inversion techniques, such as iterative Boltzmann inversion or inverse Monte Carlo,^{26,53} or more extensive modifications of the functional forms of the bonded and angle potentials in addition to retuning of the parameters. Nevertheless, the overall agreement of the coarse-grained and all-atom bonded distributions is quite good, a substantial improvement over the original Martini model, and sufficient to accurately reproduce the thermodynamic and structural properties described below.

2.1.4.2. Monomer Collapse PMFs. In Figure 2a, we present PMF curves quantifying the free energy of a monomer in solution as a function of its end-to-end distance computed from the all-atom, original Martini, and reparameterized coarse-grained models. Accurate reproduction of the all-atom PMF by a coarse-grained model is a prerequisite to realistic modeling of the equilibrium distribution between extended and collapsed configurations of a single peptide. It is apparent from the PMF curves that the original Martini model poorly reproduces the distribution of end-to-end distances observed in the all-atom simulations, whereas our coarse-grained model constructed by reparameterizing the bonded Martini interactions and incorporating backbone proper dihedrals exhibits a greatly improved agreement. Specifically, the all-atom PMF exhibits a flat-bottomed character, in which the free energy is essentially flat within error bars over the region 1.6–3.2 nm. Our coarse-grained model reproduces this feature, possessing a flat bottom over the range of 1.3–2.8 nm that is shifted relative to the all-atom model by only ~0.3 nm, marginally favoring slightly more collapsed configurations. In contrast, the original Martini model excessively favors highly collapsed peptide configurations, exhibiting a flat-bottomed minimum over the range of 0.5–1.1 nm, which is completely outside the range of that for the all-atom model. Outside of the flat-bottomed minimum, our coarse-grained model also significantly better reproduces the wings of the distribution relative to the original Martini model; however, the agreement is not quantitative. Specifically, for end-to-end distances in the range of 0.5–1.5 nm, the reparameterized coarse-grained model underpredicts the all-atom PMF, leading to an elevated preference for collapsed peptide configurations. Conversely, over the range of 1.5–3.8 nm, the coarse-grained PMF lies above that of the all-atom model, leading to a slightly suppressed preference for moderately extended conformations in the coarse-grained model; however, this trend inverts in the region above 3.8 nm such that highly extended conformations are slightly more favored. Nevertheless, the agreement between the reparameterized coarse-grained and all-atom models is very good over the flat-bottomed minimum and differs outside this region by ~5k_BT such that discrepancies in the relative stabilities of the collapsed and extended conformations predicted by the two models are within only a few multiples of the scale of thermal fluctuations.

We quantify the agreement between the reparameterized coarse-grained and all-atom PMFs using the pointwise root-mean-square distance between the optimally aligned PMF profiles

$$m_{\text{rms}} = \frac{1}{N} \sqrt{\sum_{i=1}^N (M_i^{\text{AA}} - \tilde{M}_i)^2} \quad (2)$$

where M^{AA} is the all-atom PMF and \tilde{M} is the coarse-grained PMF, optimally shifted relative to M^{AA} to minimize m_{rms} (i.e., specifying the arbitrary additive constant in the PMF to minimize the pointwise root-mean-square deviation) and linearly interpolated at the $i = 1\dots N$ points, at which M^{AA} was computed. Our reparameterized coarse-grained model better reproduces the all-atom distribution ($m_{\text{rms}}^{\text{CG}} = 0.14k_B T$ per point) compared to the unmodified Martini model ($m_{\text{rms}}^{\text{M}} = 0.24k_B T$ per point), corresponding to an average 42% improvement in the per point agreement of the PMF curves. The substantially improved agreement of our coarse-grained PMF relative to the original Martini model demonstrates the importance and value of explicit reparameterization against all-atom data. Moreover, it is notable that the reparameterized coarse-grained model uses identical nonbonded parameters to those of the original Martini model, and that this elevation in performance emerges exclusively from the reparameterization of the bonded interactions and is not due to the direct fitting of the monomer collapse PMF.

2.1.4.3. Dimerization PMFs. In Figure 2b, we present dimerization PMF curves quantifying the free energy as a function of the center-of-mass distance between a pair of monomers calculated using the all-atom, original Martini, and reparameterized coarse-grained models. In each case, we explicitly remove the trivial entropic contribution arising from restraining two objects to reside at a particular separation.⁵⁸ Reproduction of the all-atom dimerization PMF by a coarse-grained model is critical in properly modeling the strength and character of the pairwise interactions between monomers that mediate peptide assembly. The all-atom PMF reveals a relatively strong thermodynamic driving force for dimerization of $\Delta F^{\text{AA}} = (16 \pm 3)k_B T$, as measured from the plateau region at center-of-mass separations in excess of ~ 3 nm, corresponding to essentially noninteracting monomers, to the global minimum at ~ 0.25 nm, representing the dimer contact pair. The original Martini model predicts $\Delta F^{\text{M}} = (22 \pm 1)k_B T$, overestimating the driving force for dimerization by 38%. The prediction of the reparameterized coarse-grained model of $\Delta F^{\text{CG}} = (15 \pm 2)k_B T$ is in excellent agreement with the all-atom calculations. The shape of the curve along the dimerization pathway is also in quite good agreement, although the coarse-grained model slightly overestimates the all-atom PMF by a maximum of $\sim 4k_B T$ for center-of-mass separations in the range of 1.5–2.0 nm. Again, it is notable that this agreement emerges from the reparameterization of only the bonded interactions and is not due to the direct fitting of the dimerization PMF.

We again quantify the difference in the all-atom and reparameterized coarse-grained PMFs using the pointwise root-mean-square distance between the optimally aligned dimerization PMF profiles

$$d_{\text{rms}} = \sqrt{\sum_{i=1}^N (D_i^{\text{AA}} - \tilde{D}_i)^2} \quad (3)$$

where D^{AA} is the dimer all-atom profile and \tilde{D} is the dimer coarse-grained profile linearly interpolated at the $i = 1\dots N$ points at which D^{AA} was computed. In this case, the arbitrary additive constant for each PMF was specified by aligning the plateau regions of the two PMFs at large monomer separations,

corresponding to effectively noninteracting monomers. The modified coarse-grained model provides a much more accurate reproduction of the all-atom PMF ($d_{\text{rms}}^{\text{CG}} = 0.10k_B T$ per point) compared to the unmodified Martini model ($d_{\text{rms}}^{\text{M}} = 0.27k_B T$ per point), an average 62% improvement in the per point agreement of the PMF curves.

2.1.4.4. Contact Maps. In Figure 2c, we present a comparison of the contact maps computed from our all-atom and reparameterized coarse-grained models. The all-atom contact map matrix, C^{AA} , was computed as the time-averaged pairwise distances between the centers of mass of the 13 residues of the all-atom system of an isolated DFAG peptide simulated in water for 340 ns (cf. Section 2.1.1), and the coarse-grained contact map matrix, C^{CG} , was computed as the time-averaged pairwise distances between the centers of mass of the corresponding 13 residues of the coarse-grained system of an isolated DFAG peptide simulated in water for 340 ns (cf. Section 2.1.3). Such contact maps provide a time-averaged measure of the relative residue positions and therefore quantify differences in the intramolecular structure.³⁶ It is critical that a coarse-grained model accurately recapitulates the all-atom contact map to properly reproduce the peptide secondary structure. A visual comparison of the two contact maps shows a good agreement between the two models, indicating that the reparameterized coarse-grained model properly recapitulates the internal structure of the all-atom peptide. It is visually apparent that neighboring residues in the coarse-grained model tend to be slightly further separated than those in the all-atom calculations; however, no pair of residues over the entire peptide is in disagreement by more than ~ 0.5 nm. To provide a global comparative measure of pairwise residue distances in the all-atom and coarse-grained systems, we compute the Frobenius norm of the difference of the contact map matrices

$$d_{\text{cmap}}^{\text{CG}} = \sqrt{\sum_{i=1}^{r_{\text{num}}} \sum_{j=1}^{r_{\text{num}}} (c_{ij}^{\text{AA}} - c_{ij}^{\text{CG}})^2} \quad (4)$$

where $r_{\text{num}} = 13$ is the number of residues (four amino acid residues in each wing, two C-termini, and three OPV3 aromatic rings; see Figure 1), c_{ij}^{AA} is the (i, j) th element in the all-atom contact map matrix, C^{AA} , and c_{ij}^{CG} is the (i, j) th element in the coarse-grained contact map matrix, C^{CG} . The resultant value $d_{\text{cmap}}^{\text{CG}} = 1.98$ nm indicates that on average the coarse-grained contact map differs from the atomistic by only $d_{\text{cmap}}^{\text{CG}}/N^2 = 0.012$ nm/residue, indicating a globally good agreement of the secondary structure predicted by the two models.

2.1.4.5. Summary. Overall, the reparameterized coarse-grained model satisfactorily reproduces the all-atom bonded distribution functions, monomer collapse PMF, dimerization PMF, and contact maps. Although the standard Martini potential has been previously profitably employed to study the self-assembly of small peptides,³⁵ it is notable that by incorporating backbone proper dihedrals and reparameterizing the bonded interactions using all-atom data, the modified model demonstrates substantial improvements in recapitulating the dimerization PMF and secondary structure predictions that are important features governing self-assembly.³⁶ This enhancement illustrates the value of this approach in tailoring the coarse-grained model for a particular peptide chemistry to elevate the predictive performance and offers an extensible means to develop coarse-grained models for other peptide chemistries. The coarse-grained model does, however, possess two nontrivial deficiencies. First, it does not properly resolve

bimodalities in several bonds and angles observed in all-atom calculations, although this occurs for only a relatively small fraction of the bonded distribution functions and does not compromise the global structural and the thermodynamic properties assessed by the monomer and dimer PMFs and contact maps. Second, the coarse-grained model overly favors the collapsed peptide conformations relative to the all-atom calculations; however, the discrepancy is on the order of $\sim 5k_B T$ such that the effect is rather mild. In both cases, we observe that the agreement could be improved by using more sophisticated reparameterization procedures,^{26,53} modification of the functional forms of the bonded potentials, or reparameterization of the nonbonded interaction parameters. Finally, the reparameterized model offers $\sim 25\times$ speedups relative to all-atom calculations, allowing us to probe the self-assembly up to hundreds of monomers and hundreds of nanoseconds to reach time and length scales inaccessible to atomistically detailed simulations.

2.2. Coarse-Grained Molecular Dynamics Simulations of Self-Assembly. The coarse-grained model detailed in Section 2.1.2 permits us to access the time and length scales necessary to observe spontaneous peptide self-assembly in coarse-grained molecular dynamics simulations of hundreds of DFAG-OPV3-GAFD oligopeptides in water over hundreds of nanoseconds. Calculations were conducted in the Gromacs 4.6 simulation suite.⁵⁶ The initial peptide configurations were constructed from coarse-grained mappings of structures harvested from the all-atom calculations. Simulations were conducted in the NPT ensemble at 298 K and 1 bar, employing a velocity rescaling thermostat,⁵⁷ a Parrinello–Rahman barostat,⁴³ and three-dimensional periodic boundary conditions. Initial bead velocities were randomly drawn from a Maxwell–Boltzmann distribution at 298 K, and the high-energy overlaps in the initial configurations were eliminated by performing the steepest descent energy minimization to remove forces exceeding 1000 kJ/mol nm. Equations of motion were numerically integrated using a leap-frog algorithm⁴⁴ with a 5 fs time step, and the constrained bond lengths were fixed using the LINCS algorithm.⁴⁵ The Lennard-Jones interactions were shifted smoothly to zero at 1.1 nm. Electrostatic interactions were treated using reaction-field electrostatics and a relative electrostatic screening of 2.5 for polarizable water (deprotonated, high-pH model) and 15 for nonpolarizable water (protonated, low-pH model). Calculations were performed on 12 \times 2.2 GHz Intel Xeon E5 CPU cores and 2 \times NVIDIA Tesla M7020 GPUs, achieving execution speeds on the order of ~ 150 ns/day.

2.2.1. Deprotonated (High-pH) Simulations. Simulations of 343 deprotonated (-4 net charge) DFAG-OPV3-GAFD molecules were initialized by placing the oligopeptides in random locations with random orientations in an initially $33 \times 33 \times 33$ nm³ box. Charge neutrality was maintained by 1372 randomly placed counterions represented by Martini Qd beads, each carrying a +1 charge,²² and the system was solvated with Martini polarizable water⁵⁵ to achieve a peptide concentration of 15 mM. Experimentally, there is evidence that peptides do not form large aggregates under high pH conditions,^{16–19} but that they may form small aggregates, such as dimers or trimers.⁵⁹ The system was subjected to a 300 ns equilibration run to allow the cluster size distribution to attain a steady value before conducting a single 100 ns production run, over which we harvested equilibrium statistics.

2.2.2. Protonated (Low-pH) Simulations. Simulations of 42, 126, 200, 252, and 378 protonated (electrically neutral) DFAG-OPV3-GAFD molecules were conducted by insertion at random locations in random orientations into $23.5 \times 23.5 \times 23.5$ nm³ boxes and solvated with Martini nonpolarizable water⁵¹ to achieve average molar concentrations of 5.0, 15, 24, 30, and 44 mM, respectively. Each system was initially subjected to a short 50 ps run to allow the energy, temperature, and pressure to attain stable values before commencing 5 \times 400 ns independent production runs, in which we observed spontaneous and irreversible peptide aggregation. Experimentally, DFAG-OPV3-GAFD oligopeptides are known to aggregate under low-pH conditions at concentrations exceeding 10⁻³ mM¹⁸ and have been studied at concentrations up to 0.86 mM.¹⁹ We perform simulations at sixfold and higher concentrations to robustly observe the assembly of hundreds of peptides over hundreds of nanoseconds. We demonstrate below that the aggregation mechanism is independent of concentration over the range of values considered in this work.

For direct comparison with the deprotonated (high-pH) simulations, we also conducted a simulation of 343 protonated (low-pH) molecules in a $33 \times 33 \times 33$ nm³ box solvated with Martini nonpolarizable water⁵¹ at a concentration of 15 mM. It was initially subjected to a short 50 ps run to allow the energy, temperature, and pressure to attain stable values before commencing a 1.3 μ s production run.

2.2.3. Time Scale Correspondence in All-Atom and Coarse-Grained Simulations. It is well known that coarse-graining can artificially accelerate the system dynamics by eliminating degrees of freedom and smoothing the underlying free-energy landscape, and that different dynamical degrees of freedom may be accelerated by different scaling factors.^{37,51,60} As the peptide self-assembly is of the most interest to us, the relevant dynamical time scale is the center-of-mass peptide translation that governs the rate at which the molecules come into contact with one another. Accordingly, we ascertain the coarse-graining speedup by comparing the translational self-diffusion coefficients of the all-atom and coarse-grained DFAG-OPV3-GAFD molecules.^{20,61–63} Using the same simulation procedures as described in Sections 2.1.1 and 2.2.2, we conducted 10 ns simulations of an isolated peptide in a $7 \times 7 \times 7$ nm³ simulation box by employing the all-atom and coarse-grained models, respectively. We computed the self-diffusion coefficients by calculating the mean-squared displacement of the peptide center of mass and applying the Einstein relation.⁶⁴ Uncertainties were estimated by blocking the trajectory into 20 equally sized segments and computing the standard deviation of the self-diffusion coefficients calculated over these blocks. We determine a self-diffusion coefficient in the coarse-grained runs of $D_{CG} = (7 \pm 2) \times 10^{-6}$ cm²/s compared with $D_{AA} = (12 \pm 3) \times 10^{-6}$ cm²/s in the all-atom calculations. Interestingly, we observe a slight slowing of the coarse-grained dynamics, which has previously been observed for Martini representations of certain molecules including alkane chains.³⁷ Nevertheless, the self-diffusion coefficients agree within error bars, indicating that there is no significant difference in the dynamical time scales of the two simulations. Accordingly, we apply no rescaling corrections to the coarse-grained simulation time scale.

2.3. Metrics To Define a Hierarchy of Cluster Types with Different Aggregation Morphologies. Overlap and delocalization of the π orbitals in the OPV3 cores of the oligopeptides within supramolecular aggregates endow these

assemblies with emergent optoelectronic functionality.¹¹ The structural arrangement of the OPV3 cores is therefore of primary interest to characterize the observed self-assembled aggregates. Following the metrics previously developed to define clusters in asphaltene aggregation,⁶⁵ we define three criteria to determine whether or not two oligopeptides should be considered to exist in the same aggregate. The metrics, in increasing degree of strictness, are described below and illustrated schematically in Figure 3.

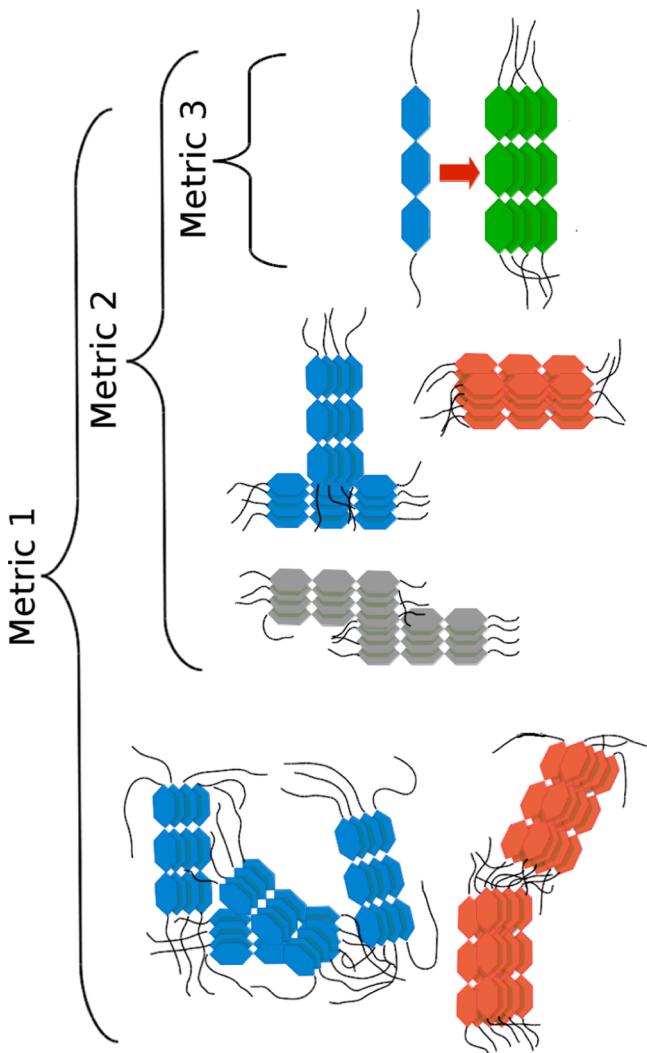


Figure 3. Schematic illustration of the hierarchy of cluster types as defined by metrics 1, 2, and 3. Metric 3 defines an “aligned cluster”, which is made up of molecules with aligned aromatic cores. Metric 2 defines an “optical cluster” composed of aligned clusters with interacting, but not necessarily aligned, aromatic cores. Metric 1 defines a “contact cluster”, which is made up of optical clusters interacting by promiscuous and nonspecific side-chain and core contacts. The distinct clusters defined by each metric are identified in different colors.

2.3.1. Metric 1: Contact Clusters. Metric 1 defines the distance between two oligopeptides a and b as

$$R_{a,b}^{(1)} = \min_{i \in a} \min_{j \in b} r_{ij} \quad (5)$$

where r_{ij} is the distance between beads i and j , and the minimization is taken over all beads in each molecule. We

define contact clusters as clusters composed of pairs of molecules that satisfy $R_{a,b}^{(1)} < r_{\text{cut}}^{(1)}$, where $r_{\text{cut}}^{(1)}$ is a tunable cutoff distance that is discussed below. Two molecules are therefore considered to be in a contact cluster if any two of their constituent atoms are within the cutoff distance of each other. Contact clusters would not necessarily be expected to demonstrate desirable optoelectronic properties, as their cores may or may not possess the $\pi-\pi$ stacking necessary for electronic delocalization.

2.3.2. Metric 2: Optical Clusters. Metric 2 defines the distance between two oligopeptides a and b as

$$R_{a,b}^{(2)} = \min_{i \in (\text{core } a)} \min_{j \in (\text{core } b)} r_{ij} \quad (6)$$

where r_{ij} is the distance between the centers of mass of aromatic rings i and j , and the minimization proceeds over the three aromatic rings within the OPV3 core of each molecule. We define optical clusters as clusters composed of pairs of molecules that satisfy $R_{a,b}^{(2)} < r_{\text{cut}}^{(2)}$. Two molecules are considered to be in an optical cluster if the minimum distance between any two aromatic rings is shorter than the cutoff distance. This metric assures proximity of the aromatic rings between the two molecules, and we term these aggregates optical clusters, as this proximity can lead to electronic delocalization and the emergence of optoelectronic properties.

2.3.3. Metric 3: Aligned Clusters. Metric 3 defines an associated distance between any two molecules a and b as

$$R_{a,b}^{(3)} = \max \left[\left(\max_{i \in (\text{core } a)} \min_{j \in (\text{core } b)} r_{ij} \right), \left(\max_{i' \in (\text{core } b)} \min_{j' \in (\text{core } a)} r_{i'j'} \right) \right] \quad (7)$$

where r_{ij} is the distance between the centers of mass of aromatic rings i and j , and the minimization and maximization are taken over three aromatic rings within the OPV3 core of each molecule. We define aligned clusters as clusters composed of pairs of molecules that satisfy $R_{a,b}^{(3)} < r_{\text{cut}}^{(3)}$. The first term in the square brackets computes the minimum distance from the center-of-mass i of each aromatic ring in molecule a to any aromatic ring center-of-mass j in molecule b and returns the largest of these distances.⁶⁵ The second term inverts the order of a and b , as the maximin operation does not commute. As previously observed,⁶⁵ this “longest shortest distance” is analogous to the mathematical definition of the diameter of a graph.⁶⁶ Accordingly, two molecules a and b are in an aligned cluster if the maximum of all minimum distances between any aromatic ring in a and any aromatic ring in b is within the cutoff distance. Pairs of molecules that meet this requirement possess well-aligned OPV3 cores, and the aggregates defined according to this metric are expected to exhibit the strongest optoelectronic properties.

2.3.4. Cutoff Distances. Following Thurston et al.,²⁰ we employed a cutoff for Metric 1 $r_{\text{cut}}^{(1)} = 0.5$ nm defined to be close to the minimum of the Lennard-Jones potential for larger atoms in the atomistic model. To determine the cutoff for Metrics 2 and 3 that explicitly considers aromatic ring interactions, we simulated a coarse-grained contact dimer of two protonated (low-pH) DFAG-OPV3-GAFD peptides in water over the course of 30 ns using the simulation parameters detailed in Section 2.2.2. We compiled histograms of the distances between the center of mass of each aromatic ring and its nearest neighbor in the other molecule comprising the dimer and specified our cutoffs as the mean of this distribution $r_{\text{cut}}^{(2)} = r_{\text{cut}}^{(3)} = 0.7$ nm.

2.3.5. Nesting of Metrics. We observe that the metrics are nested such that a pair of peptides identified as an aligned cluster (Metric 3) will also be identified as an optical cluster (Metric 2) and a contact cluster (Metric 1), and that a pair identified as an optical cluster is also identified as a contact cluster. The inverse, however, is not true. A pair of peptides identified as a contact cluster may or may not be an optical or aligned cluster. Accordingly, for any group of peptides, the contact cluster size is constrained to be equal to or greater than the optical cluster size, which in turn is constrained to be equal to or greater than the aligned cluster size.

3. RESULTS AND DISCUSSION

3.1. Morphological Characterization of Hierarchical Self-Assembly. We have constructed a coarse-grained model of DFAG-OPV3-GAFD π -conjugated oligopeptides explicitly parameterized against the all-atom simulation data. This efficient model allows us to directly simulate the spontaneous self-assembly of hundreds of peptides over hundreds of nanoseconds, allowing us to characterize the molecular-level morphology and kinetics of the early-stage assembly of optically active aggregates.

3.1.1. Assembly of Deprotonated (High-pH) Molecules Is Self-Limiting. It is known from experiment that deprotonated (high-pH) DFAG-OPV3-GAFD molecules do not assemble large aggregates due to electrostatic repulsion between their negatively charged Asp termini;^{16,17,19} however, there is recent experimental⁵⁹ and computational²⁰ evidence that the formation of small aggregates is thermodynamically spontaneous due to favorable dispersion, π -stacking, and hydrophobic interactions. Our simulations of deprotonated peptides at 15 mM corroborate this result, showing that the cluster size distributions under each of the three cluster metrics (cf. Section 2.3) attain stable equilibrium distributions within 300 ns of simulation (Figure 4a). This observation corresponds to the self-limiting aggregation of peptides into small aligned and optical clusters containing up to five monomers and small contact clusters containing up to eight. The high concentration of negative charge within these small aggregates supplied by the deprotonated Asp termini prohibits further assembly into larger aggregates.

3.1.2. Assembly of Protonated (Low-pH) Molecules Is Hierarchical and Irreversible. We present in Figure 5 the mass-averaged cluster size of the contact, optical, and aligned clusters as a function of time for each of the five concentrations considered under protonated (low-pH) conditions. Contact clusters (Metric 1) exhibit irreversible aggregation that leads to the formation of increasingly large aggregates over the course of our 400 ns production runs. Assembly initially occurs by the aggregation of monomers and then proceeds by the agglomeration of peptide aggregates and proceeds more rapidly at higher concentrations. In Figure 4b, we present the time-resolved contact cluster size distribution at 15 mM. These observations are consistent with experimental data showing the irreversible assembly of up to micron-sized aggregates under low-pH conditions.^{16–18,59}

Optical clusters (Metric 2) also exhibit irreversible and unbounded aggregation. At 5.0 mM, we observe little difference in the average size of contact and optical clusters, which indicates that optical clusters are not associating into larger contact clusters through side-chain interactions over the 400 ns simulation time scales. At higher concentrations, the optical clusters are substantially smaller than the contact clusters,

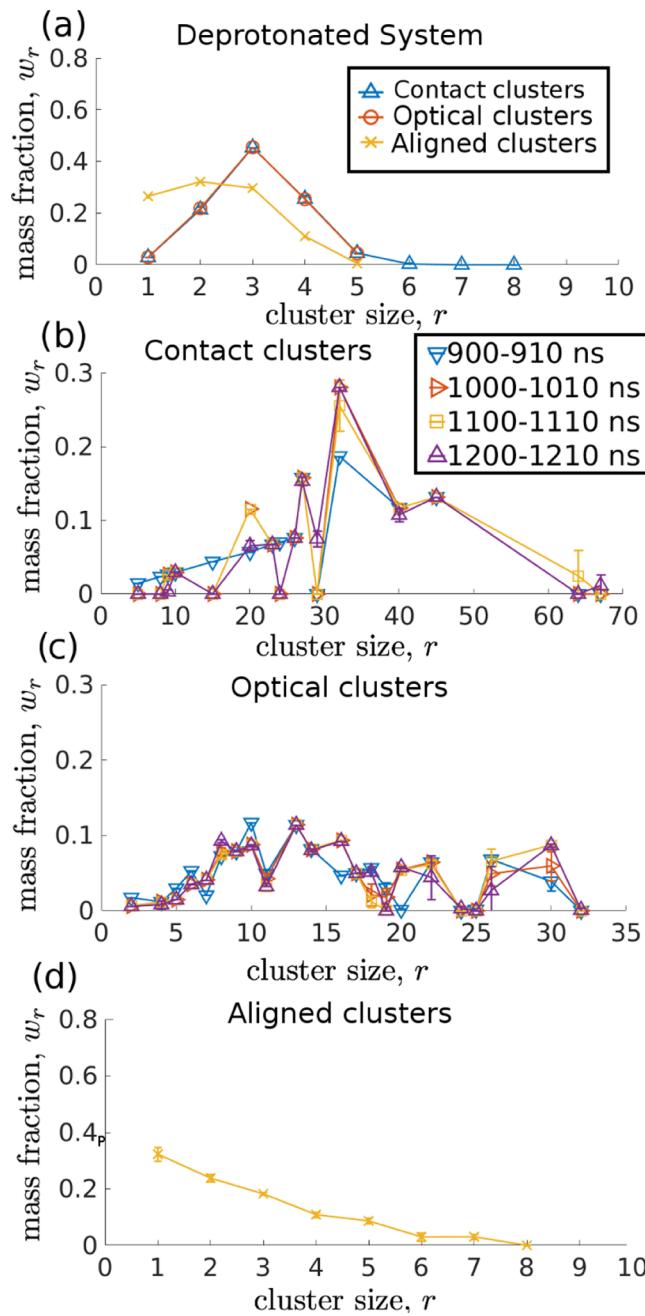


Figure 4. Cluster size distributions in the deprotonated (high-pH) and protonated (low-pH) systems at 15 mM. (a) Mass fraction (w_r) of peptides in the deprotonated (high-pH) system that exist in clusters of size r defined according to the three metrics. The mass fraction distributions under all three metrics cease to evolve within the 300 ns equilibration period simulation time, indicating that the aggregation process has attained a dynamic equilibrium, and data are averaged over the 100 ns production run. In the protonated (low-pH) state, the mass fraction distributions for aligned clusters reach steady state after \sim 200 ns; however, the optical and contact cluster distributions never reach a terminal distribution and continue to shift to larger cluster sizes over the course of the entire 1.3 μ s production run, indicating unbounded and irreversible aggregation, although the process is rather slow at this concentration. (b) Mass fraction distribution of protonated (low-pH) contact clusters averaged over the four time windows: 900–910, 1000–1010, 1100–1110, and 1200–1210 ns. (c) Mass fraction distribution of protonated (low-pH) optical clusters averaged over the four time windows, 900–910, 1000–1010, 1100–1110, and 1200–1210 ns, following the legend in (b). (d) Steady-state mass fraction

Figure 4. continued

distribution of aligned clusters in the protonated (low-pH) state averaged over the terminal 100 ns of the production run. Uncertainties in all cases correspond to standard errors estimated by block averaging. Error bars not explicitly displayed are smaller than the symbol size. Lines connecting the points are provided to guide the eye.

indicating the presence of a higher-order association through side-chain interactions. We observe an initially high rate of formation of optical clusters that slows at ~ 200 ns and an average cluster size of approximately 5–10 molecules, which indicates a cluster-size-dependent mechanism of aggregation. Larger optical clusters are both less mobile and have less favorable surface area through which to form new optical clusters. Optical clusters are also formed by rearrangements or “ripening” of contact clusters to result in more favorable configurations. In Figure 4c, we present the time-resolved optical cluster size distribution at 15 mM.

In contrast to the unbounded growth of contact and optical clusters, aligned clusters (Metric 3) exhibit self-limiting behavior. After ~ 200 ns for all concentrations studied, we observe the emergence of a mass-averaged cluster size of aligned clusters of $\mu_2^{\text{aligned}} \approx 2$. In Figure 4d, we present the aligned cluster size distribution at 15 mM, which reveals a decreasing distribution in aligned cluster sizes up to a maximum size of 8. The most probable aligned cluster size is a monomer, with 30% of peptides existing in this state. Dimers, trimers, and tetramers account for 25, 22, and 11% of peptides, respectively, leaving only 12% as higher-order aligned aggregates.

Our results reveal a hierarchical assembly mechanism at low pH, wherein peptides rapidly achieve a steady-state distribution of aligned clusters composed of ordered stacking between all three aromatic rings in the OPV3 cores. This distribution is dominated by small clusters, with 88% of peptides in aligned clusters existing as monomers, dimers, trimers, or tetramers. These aligned clusters form the components of larger optical clusters formed from proximate interactions between the

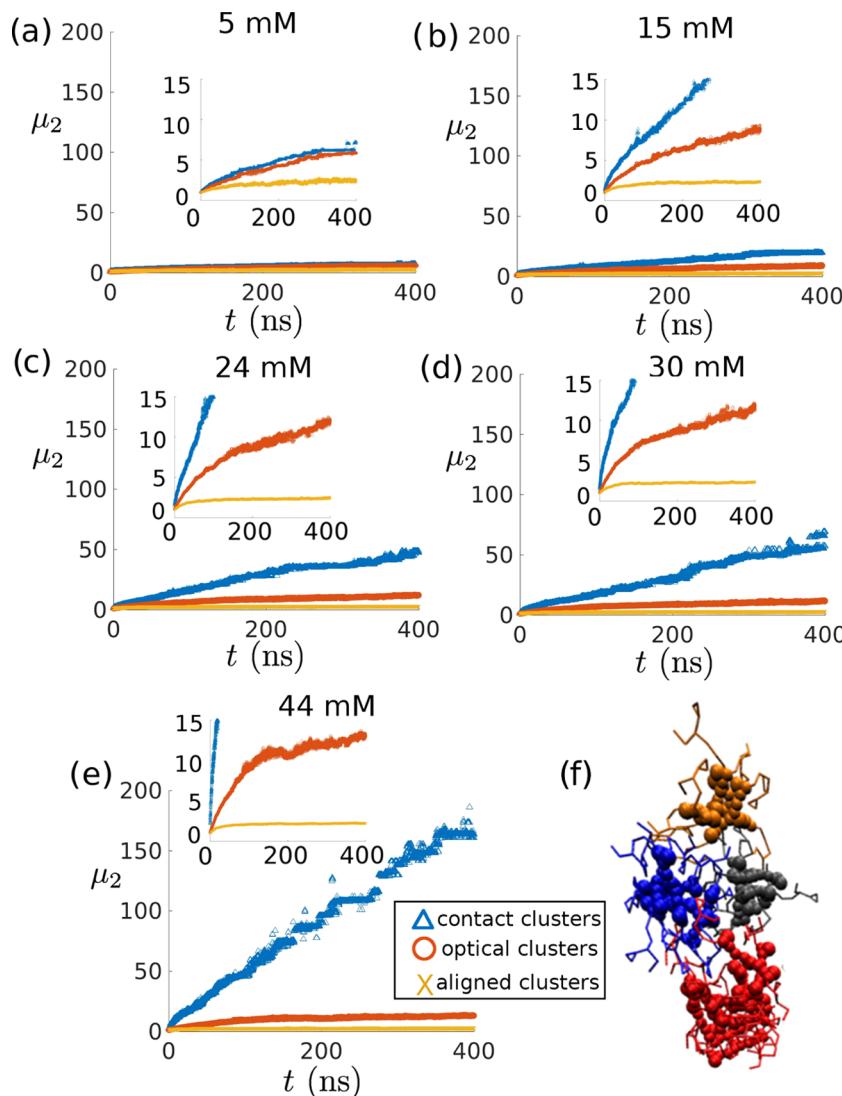


Figure 5. Mass-averaged cluster size as a function of time according to the three cluster metrics for the protonated (low-pH) system at concentrations of (a) 5.0 mM, (b) 15 mM, (c) 24 mM, (d) 30 mM, and (e) 44 mM, where all panels follow the legend displayed in panel (e). Systems were initialized from randomly oriented and randomly placed monomers. The insets show a zoom to better resolve the time evolution for the optical and aligned clusters. (f) A representative snapshot of a contact cluster, in which the constituent optical clusters are distinguished by different colors. All molecular visualizations in this work were constructed using VMD.⁶⁷

aromatic cores, but lacking the in-register stacking and alignment of the aligned clusters. In contrast to the aligned clusters, the optical clusters do not attain a steady-state cluster size distribution but continue to increase in size over the entire simulation, by both agglomeration with other optical clusters and internal rearrangements, in which portions of contact clusters with non-core interactions ripen into optical clusters through improved ordering and stacking of the constituent monomers. Beyond ~ 200 ns, this growth is relatively slow, and the mass-averaged optical cluster size in the time window 200–400 ns is $\mu_2^{\text{optical}} \approx 10$. These optical clusters, in turn, serve as the building blocks for still larger contact clusters that are formed by promiscuous and nonspecific contacts between optical cluster side chains and cores. Contact clusters also grow without bound, assembling by irreversible agglomeration to ultimately subsume all peptides in the system. We present in Figure 6 representative snapshots of the latter stages of assembly at each of the concentrations studied.

3.1.3. Contact Clusters Are Fractal Aggregates. To characterize the morphology of the self-assembling contact clusters of protonated (low-pH) peptides defined according to Metric 1, we estimate the fractal dimensionality of the system at different scales by computing the correlation dimension. The correlation dimension, D , is a measure of the space-filling nature of a system, which is estimated from the correlation integral $C(r)$.⁶⁸ For a set of N points in space, the correlation integral is defined as

$$C(r) = \lim_{N \rightarrow \infty} \frac{g}{N^2} \quad (8)$$

where g is the number of pairs of points separated by a distance less than r .⁶⁸ The fractal dimension is related to the correlation integral through the relationship

$$C(r) \sim r^D \quad (9)$$

which provides a measure of how rapidly new points are accumulated into an expanding sphere around a tagged point. We numerically estimate the correlation integral for our system as^{65,69}

$$C(r) = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^N H(r - r_{ij}^{\text{COM}}) \quad (10)$$

where r_{ij}^{COM} is the distance between the centers of mass of monomers i and j and $H(x)$ is the Heaviside step function

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad (11)$$

In Figure 7a–e, we present the log–log plot of $C(r)$ versus r for the final snapshot of the five independent runs conducted at each concentration. In each case, we observe approximately three regimes: the first spanning approximately 0.13–1.0 nm, the second approximately 1.0–7.4 nm, and the third approximately 7.4–20 nm. The shortest length scale regime approximately corresponds to the characteristic size of aligned and optical clusters in the system (cf. Figure 9a), the medium length scale regime to smaller contact clusters containing approximately three to five optical clusters (cf. Figure 9b), and the longest length scale regime to large contact clusters, which saturate at large r due to the finite size of the simulation box (cf. Figure 9c). At the lowest concentrations, that is, 5 and 15 mM, the second regime comprises largely the interstitial solvent

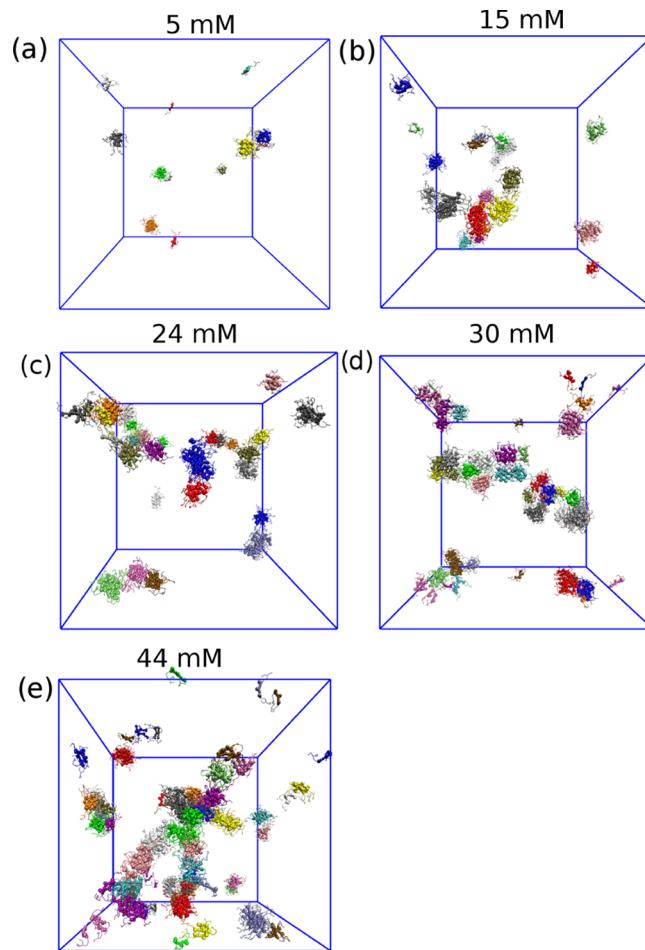


Figure 6. Representative snapshots of the late-stage (>200 ns) assembly of the protonated (low-pH) peptides at concentrations of (a) 5.0 mM, (b) 15 mM, (c) 24 mM, (d) 30 mM, and (e) 44 mM. At this time, the aligned cluster size distribution has attained steady state. The optical cluster size distribution continues to move toward larger clusters relatively slowly through agglomeration and internal rearrangements of contact clusters. The contact clusters are formed by unbounded and irreversible agglomeration, and by this time, most of the aggregation in the system occurs through agglomeration of optical clusters containing >10 peptides. For ease of viewing, in each panel, we have indicated the distinct optical clusters in different colors, where core beads are represented as spheres, side chains as lines, and water has been removed for clarity.

between isolated aggregates as demonstrated by the flat nature of $C(r)$, and the third regime characterizes the characteristic separations between these clusters (cf. Figure 6a,b). At higher concentrations, the peptides begin to form a connected network and $C(r)$ possesses a large positive slope in all three regimes (cf. Figure 6c–e).

In Figure 7f, we present the fractal dimension extracted from linear fits constructed within each of the three regions as a function of concentration. In the first region, the fractal dimension is ~ 2 for all concentrations studied, corresponding to the growth of small aligned and optical clusters into slightly elongated ellipses (cf. Figure 9a). The second region has a fractal dimension that grows with concentration, from ~ 0.5 to ~ 1.5 , corresponding to the departure from isolated aligned and optical clusters to the formation of increasingly large contact clusters. The third region has a fractal dimension of ~ 1.5 for all

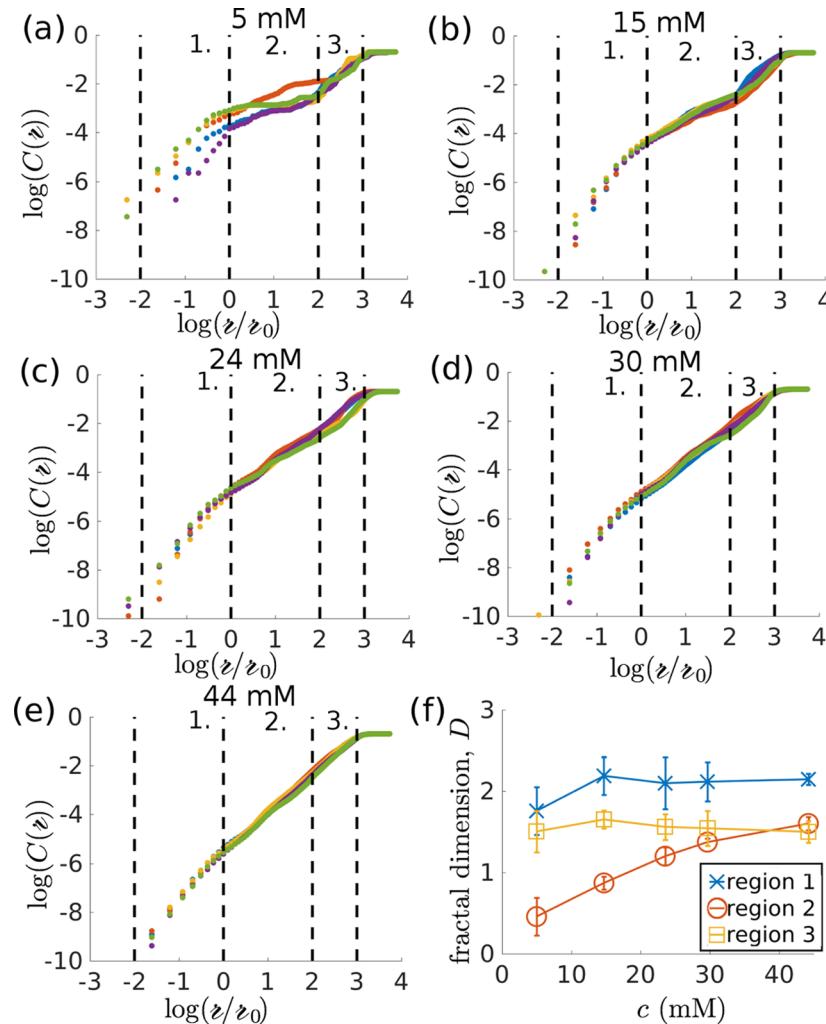


Figure 7. Fractal dimension of the hierarchical aggregates as a function of length scale. (a–e) Log–log plots of correlation integral $C(r)$ vs radius r at each concentration. We demarcate the three regimes that emerge from this analysis by dashed black lines. We render r dimensionless before taking the logarithm by rescaling by an arbitrary reference distance $r_0 = 1$ nm. (f) Fractal dimension D computed from linear fits to each of the three regimes as a function of concentration.

concentrations, reflecting the formation of a fractal network of branched contact clusters (cf. Figure 9c).

3.1.4. Contact Clusters Assemble Asymmetrically into Elongated Ellipsoidal Aggregates That Display a Branching Behavior. To further characterize and explore the three structural regimes identified in the analysis of the fractal dimension, we computed the gyration tensor, S , for each contact cluster defined according to Metric 1 observed in our simulations as⁷⁰

$$\mathbf{S}_{pq} = \frac{1}{2R^2} \sum_{i=1}^R \sum_{j=1}^R (r_p^{(i)} - r_p^{(j)}) (r_q^{(i)} - r_q^{(j)}) \quad (12)$$

where p and q index over the three Cartesian coordinates $\{x, y, z\}$, i and j index over the beads of all peptides in the cluster, $R = 29r$ is the total number of beads in the cluster, r is the number of molecules in the cluster, and there are 29 coarse-grained beads in each molecule (cf. Figure 1). The eigenvalues of S correspond to the squared ordered principal moments of the cluster $\{\xi_i(r)\}_{i=1}^3$ that characterize its extent along its principal axes furnished by the corresponding eigenvectors.⁷⁰ The principal moments are related to the radius of gyration as

$R_g = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}$.⁷⁰ In performing these calculations, we eliminate any potentially spurious effects introduced by the periodic walls by excising the cluster from the simulation box and unwrapping the periodic boundary conditions before computation of the gyration tensor. This procedure will fail for formally infinite aggregates that span the fundamental simulation cell; however, we did not observe any such cluster at the concentrations studied in this work.

We plot in Figure 8a the ensemble-averaged radius of gyration $\langle R_g(r) \rangle$ and principal moments of the gyration tensor $\{\langle \xi_i(r) \rangle\}_{i=1}^3$ as a function of cluster size r averaged over all clusters observed in our simulations. As expected, $\langle R_g(r) \rangle$ and $\{\langle \xi_i(r) \rangle\}_{i=1}^3$ all trend upward with increasing r . We plot in Figure 8b the ensemble-averaged ratios of the principal moments, $\{\langle \xi_i / \xi_j \rangle\}_{j>i}$. Ratios close to unity characterize isotropic aggregates, whereas values far from unity are indicative of high-aspect-ratio clusters. In Figure 8c, we plot the ensemble-averaged relative shape anisotropy, $\langle \kappa^2(r) \rangle$, as a function of cluster size, defined as⁷⁰

$$\kappa^2 = \frac{3}{2} \frac{\xi_1^4 + \xi_2^4 + \xi_3^4}{(\xi_1^2 + \xi_2^2 + \xi_3^2)^2} - \frac{1}{2} \quad (13)$$

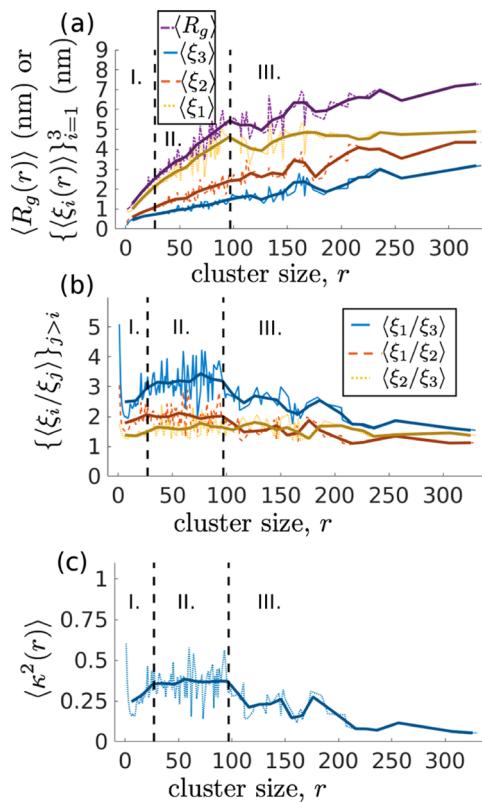


Figure 8. Gyration tensor analysis of contact clusters formed by protonated (low-pH) peptides as a function of cluster size r averaged over clusters observed in our simulations at all times and concentrations. (a) Ensemble-averaged radius of gyration $\langle R_g(r) \rangle$ and principal moments of the gyration tensor $\{\langle \xi_i(r) \rangle\}_{i=1}^3$. (b) Ensemble-averaged ratios of the principal moments of the gyration tensor $\{\langle \xi_i/\xi_j \rangle\}_{j>i}$. (c) Ensemble-averaged relative shape anisotropy $\langle \kappa^2(r) \rangle$. The thin lines correspond to mean values computed over all clusters observed in our simulations at all concentrations, and the heavy lines denote sliding window smoothing of the data using a window size of 10 to more clearly show the gross trends in the data. The three aggregation regimes—Regime I ($1 \leq r < 27$), Regime II ($27 \leq r < 97$), and Regime III ($r \geq 97$)—are indicated by dashed vertical lines.

which is bounded to the range $[0, 1]$, where $\kappa^2 = 0$ indicates that the beads in the cluster are distributed with tetrahedral or higher-order symmetry, and $\kappa^2 = 1$ indicates that they are arranged along a line.

The plots of $\{\langle \xi_i/\xi_j \rangle\}_{j>i}$ and $\kappa^2(r)$ again inform a division into three regimes consistent with those identified in the fractal dimension analysis in Section 3.1.3, but this time partitioned according to cluster size r rather than the observation length scale: Regime I ($1 \leq r < 27$), Regime II ($27 \leq r < 97$), and Regime III ($r \geq 97$). We present representative cluster snapshots from each regime in Figure 9. In Regime I ($1 \leq r < 27$), we observe an initially rapid decrease in all three principal moment ratios and a concomitant decrease in the relative shape anisotropy, $\langle \kappa^2(r) \rangle$, from 0.6 to 0.2 over the cluster size range of $1 \leq r < 6$. This is followed by the recovery of $\langle \kappa^2(r) \rangle$ from 0.2 to 0.4 and $\langle \xi_1(r)/\xi_3(r) \rangle$ from 2 to 3 over the range of $6 \leq r < 27$, whereas $\langle \xi_1(r)/\xi_2(r) \rangle$ and $\langle \xi_2(r)/\xi_3(r) \rangle$ remain approximately flat at 2 and 1.5, respectively. Structurally, this corresponds to the aggregation of elongated high-aspect-ratio monomers into more spherical optical clusters comprising several monomers, followed by the aggregation of

these spherical clusters into more elongated, ellipsoidal optical and contact clusters (cf. Figure 9a). In Regime II ($27 \leq r < 97$), the three principal moment ratios and relative shape anisotropy remain approximately constant. At these sizes, the clusters no longer aggregate in a preferential direction but begin to isotropically branch, preserving the relative aspect ratios with the increase of the cluster size (cf. Figure 9b). In Regime III ($r \geq 97$, up to the maximum cluster size of $r = 330$ observed in our simulations), $\langle \xi_1(r)/\xi_3(r) \rangle$ declines from 3 to 1.5, $\langle \xi_1(r)/\xi_2(r) \rangle$ declines from 2 to 1, and $\langle \xi_2(r)/\xi_3(r) \rangle$ remains at 1.5, with a corresponding drop in $\langle \kappa^2(r) \rangle$ from 0.4 to 0.1. This aggregation regime corresponds to the continued formation of a large-scale self-similar branched network lacking any preferred directionality and with a fractal dimension of ~ 1.5 (cf. Figure 9c).

3.2. Smoluchowski Coagulation Model Describes the Kinetics of Contact Cluster Aggregation. Our simulations of the protonated (low-pH) peptides provide information about the kinetics of the hierarchical peptide self-assembly over hundreds of nanoseconds. By mining these data to parameterize a kinetic model of assembly, we can both garner insight into the microscopic assembly dynamics and extrapolate our predictions to time and length scales beyond the reach of all-atom or coarse-grained molecular simulation.

The Smoluchowski coagulation model is a general theory describing irreversible aggregation in well-mixed systems^{71–73} that has been applied to a number of different systems, including planetary aggregation,^{74,75} aerosol droplet coagulation,⁷⁶ and amyloid protofibril formation.⁷⁷ This model assumes that aggregation is irreversible (i.e., fragmentation is forbidden), the state of the system is fully specified by the cluster size distribution without differentiating clusters with regard to their particular morphologies, and the system can be described as well mixed so that clusters need not be spatially resolved.

We propose that the Smoluchowski coagulation model can provide a quantitative description of the DFAG-OPV3-GAFD contact cluster aggregation kinetics. It is an appropriate model for this process for several reasons. First, we observe the aggregation of these oligopeptides under low-pH conditions to be effectively irreversible over several hundred nanoseconds (cf. Figure 5), making the explicit treatment of fragmentation unnecessary. Second, we do not observe any lag phase in either simulation or experiment,^{17,19,20} suggesting that the aggregation proceeds downhill in free energy and that nucleation-dependent models are unlikely to provide a good description.⁷⁸ Third, although we have previously used Markov state models to treat the short-time aggregation dynamics,²⁰ Smoluchowski coagulation offers a more physically motivated model that can be straightforwardly extrapolated to longer length and time scales. Fourth, we initialize our systems as dispersed monomers meaning that the well-mixed assumption is expected to be a good description over the time scales of our simulations and is consistent with the absence of spatial gradients in experiment.¹⁷ Fifth, any dependence of the aggregation kinetics of the observed morphological changes in the growing clusters can, if necessary, be straightforwardly captured within the Smoluchowski model as a cluster-size-dependent rate constant.

3.2.1. Smoluchowski Coagulation Model. The continuous time Smoluchowski coagulation equation for systems admitting discrete aggregate sizes is⁷³

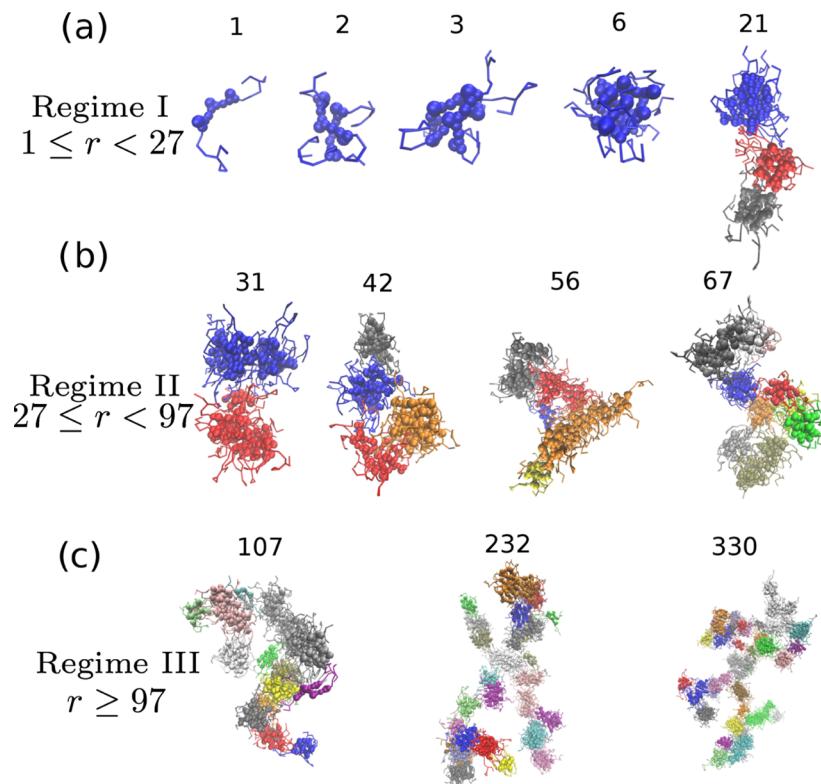


Figure 9. Representative contact clusters illustrating the hierarchical assembly process in each of the three aggregation regimes: (a) Regime I ($1 \leq r < 27$), (b) Regime II ($27 \leq r < 97$), and (c) Regime III ($r \geq 97$). Distinct optical clusters within the contact clusters are distinguished by different colors, and the number of monomers in each image are indicated by the accompanying numeral. Each panel is scaled differently for ease of viewing the clusters of disparate sizes; however, in all cases, the end-to-end distance of a fully extended monomer is ~ 3.5 nm, and the length of each OPV3 core is ~ 1.5 nm.

$$\frac{dn_r}{dt} = \frac{1}{2} \sum_{i=1}^{r-1} K_{i,r-i} n_i(t) n_{r-i}(t) - \sum_{i=1}^{\infty} K_{r,i} n_r(t) n_i(t) \quad (14)$$

where t is the time, K_{ij} is a kernel corresponding to a second-order rate constant for coagulation between aggregates of size i and size j , and $n_r(t)$ is the number concentration of aggregates of size r at time t . Analytical expressions for $n_r(t)$ are unavailable for arbitrary kernels, but solutions are known for the special cases of (i) constant $K_{ij} = K$,^{73,79} (ii) additive $K_{ij} = K(i + j)$,⁷⁹ and (iii) multiplicative $K_{ij} = K(i)K(j)$ kernels.⁸⁰ For the homogeneous kernel $K_{ai,aj} = a^2 K_{i,j}$ expressions for $n_r(t)$ are not known, but the moments and their ratios, such as the number-averaged and mass-averaged cluster sizes, are available.^{72,77}

As we demonstrate below, the simplest assumption of a constant kernel $K_{ij} = K$, corresponding to size-independent coagulation rate constants, provides good fits to our simulation data. Under this assumption, the solution for $n_r(t)$ under arbitrary initial conditions is⁷⁹

$$n_r(t) = 4 \frac{n_r(0)}{(KtM_0(0) + 2)^2} + \left(\frac{KtM_0(0)}{KtM_0(0) + 2} \right) \sum_{j=1}^{r-1} \frac{n_{r-j}(0) n_j(t)}{M_0(0)} \quad (15)$$

where $n_r(0)$ is the initial number concentration of r -mers and $M_i(t) = \sum_{r=1}^{\infty} r^i n_r(t)$ is the i th moment of the cluster size distribution at time t so that $M_0(0) = \sum_{r=1}^{\infty} n_r(0)$ is the sum of the initial number concentrations of all r -mers. For monodisperse initial conditions, this simplifies to

$$n_r(t) = 4 \frac{M_1}{(KtM_1 + 2)^2} \left(\frac{KM_1 t}{KM_1 + 2} \right)^{r-1} \quad (16)$$

where $M_1(t) = \sum_{r=1}^{\infty} r n_r(t) = M_1(0) = M_1$ is the total concentration of monomers in solution, which is related to the concentration through Avogadro's number as $c(t) = M_1(t)/N_A$. The number-averaged cluster size, $\mu_1(t)$, under arbitrary initial conditions is^{73,79}

$$\mu_1(t) = \frac{M_1(t)}{M_0(t)} = \mu_1(0) + \frac{1}{2} KM_1 t \quad (17)$$

and the mass-averaged cluster size, $\mu_2(t)$, under arbitrary initial conditions is^{73,77,79}

$$\mu_2(t) = \frac{M_2(t)}{M_1(t)} = \mu_2(0) + KM_1 t \quad (18)$$

and following Modler et al.,⁷⁷ we identify

$$t_c = \frac{2}{KM_1} \quad (19)$$

as a characteristic coagulation time scale, where it is conventional to retain the factor of 2 in the numerator.

3.2.2. Understanding the Success of the Constant Smoluchowski Kernel. Assuming the constant kernel to be of the form $K_{ij} = K = \alpha k_s$, we follow Modler et al.⁷⁷ to decompose the coagulation rate constants into the product of the sticking probability, α , accounting for the probability of a productive collision, and an intrinsic rate constant, k_s . Assuming diffusion-

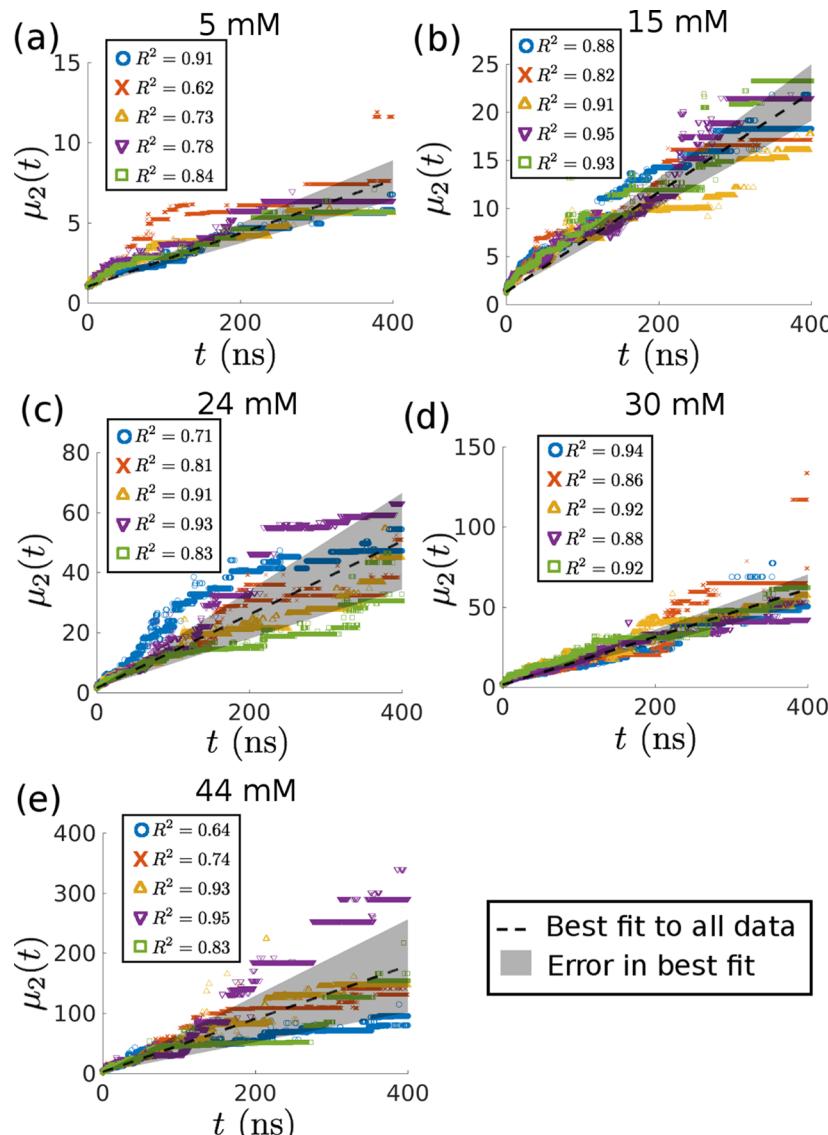


Figure 10. Mass-averaged cluster size as a function of time $\mu_2(t)$ measured in simulations at concentrations of (a) 5.0 mM, (b) 15 mM, (c) 24 mM, (d) 30 mM, and (e) 44 mM. The black lines indicate the least-squares linear fits to all five independent runs. The shaded area indicates the uncertainty in the best fit computed from the standard deviation in the slopes computed for each individual run. Data from each independent simulation are presented in different colors using different symbols and labeled with the R^2 value of the best-fit line.

controlled aggregation, we can develop an approximate expression for the intrinsic rate constant by inserting the Stokes–Einstein relation for the diffusivity of a spherical particle of radius R through a low Reynolds number liquid

$$D = \frac{k_B T}{6\pi\eta R} \quad (20)$$

into the analytical expression of the rate constant for the collision of two Brownian hard spheres of radii R_1 and R_2

$$k_s = 4\pi(R_1 + R_2)(D_1 + D_2) \quad (21)$$

to yield^{77,81,82}

$$k_s = \frac{2k_B T}{3\eta} \left(2 + \frac{R_1}{R_2} + \frac{R_2}{R_1} \right) \quad (22)$$

where T is the temperature, k_B is Boltzmann's constant, and η is the viscosity of the solvent. Equation 22 provides a means to understand the success of the constant kernel in recapitulating

our simulation data. First, we showed in Section 3.1.3 that contact clusters are fractal objects of dimensionality $D \approx 1.5$, meaning that their characteristic radius R scales weakly with cluster size r as $R \sim r^{2/3}$. Second, although the time-resolved contact cluster size distributions in Figure 4b are relatively broad, they span no more than an order of magnitude over the course of our simulation, meaning that at any time instant we are unlikely to observe collisions between clusters of vastly discrepant sizes. Accordingly, the terms $\frac{R_i}{R_j} \sim \left(\frac{r_i}{r_j}\right)^{2/3}$ are both expected to be of order unity, leading us to predict an approximately size-independent intrinsic rate constant

$$k_s \approx \frac{8k_B T}{3\eta} \quad (23)$$

Assuming that the sticking probability, α , is not a strong function of cluster size, this analysis provides a theoretical rationalization for the use of a size-independent Smoluchowski

coagulation kernel. Modler et al.⁷⁷ have previously demonstrated the validity of this assumption in the context of amyloid peptides by showing a good agreement between numerical solutions employing size-dependent and size-independent intrinsic rate constants.

3.2.3. Inference of the Smoluchowski Coagulation Rate Constant. We now analyze our coarse-grained molecular simulation trajectories to determine the best-fit constant kernel, K , for the Smoluchowski coagulation model. Appealing to eq 18, we plot the mass-averaged contact cluster size measured in our simulations as a function of time $\mu_2(t)$ and compute the least-squares linear fit, from which we estimate K by dividing the slope by the total monomer concentration, M_1 . The linear fits at each concentration are presented in Figure 10, and the best-fit values of K are presented in Table 1.

Table 1. Coagulation Rate Constant K as a Function of Concentration, Estimated from the Least-Squares Linear Fits of the Measured Mass-Averaged Contact Cluster Size $\mu_2(t)$ Using eq 18^a

concentration, c (mM)	K (ns ⁻¹ nm ³)
5.0	6 ± 1
15	5.9 ± 0.8
24	9 ± 3
30	8 ± 1
44	17 ± 7

^aUncertainties are estimated from the standard deviation, σ_K in K computed over the five independent production runs.

Inserting the best-fit coagulation rate constant into the analytical Smoluchowski expressions for $n_r(t)$ (eq 15) shows good agreement with the number concentration time courses measured in the simulation. In Figure 11, we compare the simulation results and the analytical expressions for the number concentrations from monomers to 12-mers in the 24 mM system and provide analogous plots for the other concentrations in Figures S6–S9. The agreement for larger aggregates deteriorates somewhat because of both a paucity of simulation data due to the relative rarity of large aggregates and finite size effects, wherein the finite number of monomers in the simulations stands in contrast to the assumption of the Smoluchowski model of an infinite system and monomer supply (Figure S10). Nevertheless, this simple spatially unresolved model assuming size-invariant coagulation rate constants is capable of quantitatively recapitulating the number fraction time courses of the small to medium-sized clusters for which we have good simulation data, which supports the Smoluchowski coagulation model as a good description of the aggregation kinetics.

3.2.4. Characteristic Coagulation Time Has a Simple Concentration Dependence. The characteristic coagulation time is inversely proportional to both the coagulation rate constant and the monomer concentration, $t_c = \frac{2}{KM_1} = \frac{2}{KcN_A}$ (eq 19). Employing the best-fit values of K in Table 1, the plot of t_c as a function of c in Figure 12 reveals an approximate power law dependence of the form

$$\log(t_c/t_c^0) = -\gamma \log(c/c^0) + \log(\zeta) \quad (24)$$

where $t_c^0 = 1$ ns and $c^0 = 1$ mM are, respectively, the arbitrary reference time scale and concentration to make the arguments of the logarithm dimensionless. Linear least-squares regression,

in which we handled the observed heteroskedasticity in the data by weighting each data point in inverse proportion to its variance,⁸³ reveal the best-fit parameters of $\gamma = 1.3$ (95% CI: 0.9, 1.7) and $\zeta = 1000$ (95% CI: 400, 3000). A concentration invariant coagulation rate constant would produce a simple $t_c \propto c^{-1}$ dependence, and the observed deviation from this scaling reflects the concentration dependence of K (Table 1).

At peptide concentrations of 0.86 mM (0.1 mg/mL) realized in the directed assembly of peptide oligomers in microfluidic devices,¹⁹ our best-fit parameters predict a coagulation time $t_c = 800$ ns (95% CI: 200 ns, 1400 ns) in good agreement within error bars of the 2500 ns (95% CI: 500 ns, 9300 ns) characteristic assembly time scale previously predicted under these conditions by Thurston et al.²⁰ using Markov state models fitted to all-atom implicit solvent calculations.

Finally, we observe the powerful interpretation of Figure 12 as a calibration curve relating the concentration, c , as a macroscopic experimental control parameter to the microscopic characteristic coagulation time, t_c . This relationship can be used to infer, for a particular concentration, the characteristic coagulation time, t_c , and therefore the microscopic Smoluchowski model coagulation rate constant through $K = \frac{2}{t_c c N_A}$.

The estimate of K provides access to predictions of the microscopic temporal evolution of each cluster size $n_r(t)$ and moments of the cluster size distribution through the analytical solutions in eqs 15–19.

3.2.5. Cluster Size Evolutions Exhibit Approximate Scaling Collapse. Under a homogeneous coagulation kernel $K_{ai,aj} = a^\lambda K_{ij}$, of which the constant kernel is a special case with $\lambda = 0$, the Smoluchowski model predicts that if $n_r(t)$ is a solution to eq 14, then so is $\frac{a^{j+1}}{b} n_{ar}(bt)$. This self-similarity leads to the prediction of a scaling form of the solution valid for long times and large clusters^{72,73,76,84}

$$n_r(t) \sim s(t)^{-2} \phi(r/s(t)) \quad (25)$$

where $\phi(x)$ is a master function that is the same for all $n_r(t)$, and $s(t)$ is the ratio of any two consecutive moments of the cluster size distribution⁷⁶

$$s(t) = \frac{M_i(t)}{M_{i-1}(t)} \quad (26)$$

This “scaling hypothesis” states that for long times and large clusters, all number concentrations will evolve to a self-similar distribution of the form of eq 25, regardless of the precise initial conditions, as long as those initial conditions are sufficiently narrow (i.e., not power-law-distributed). Although the hypothesis remains unproven for arbitrary kernels and initial conditions, it has been supported by substantial numerical evidence and is proven for the particular case of a constant kernel under monodisperse initial conditions of interest here.^{72,73,76,85}

Adopting $s(t) = \frac{M_2(t)}{M_1(t)} = \mu_2(t)$, if Smoluchowski coagulation serves as a good model for peptide aggregation, we should expect to see scaling collapse in our data of the form

$$n_r(t)\mu_2(t)^2 \sim \phi(r/\mu_2(t)) \quad (27)$$

We plot $n_r(t)\mu_2(t)^2$ against $(r/\mu_2(t))$ for all contact clusters of size greater than 20 at each concentration in Figure 13 and for contact clusters of size 1, 3, 5, 11, 20, and 30 aggregated over all concentrations studied in Figure 14. We exclude clusters with

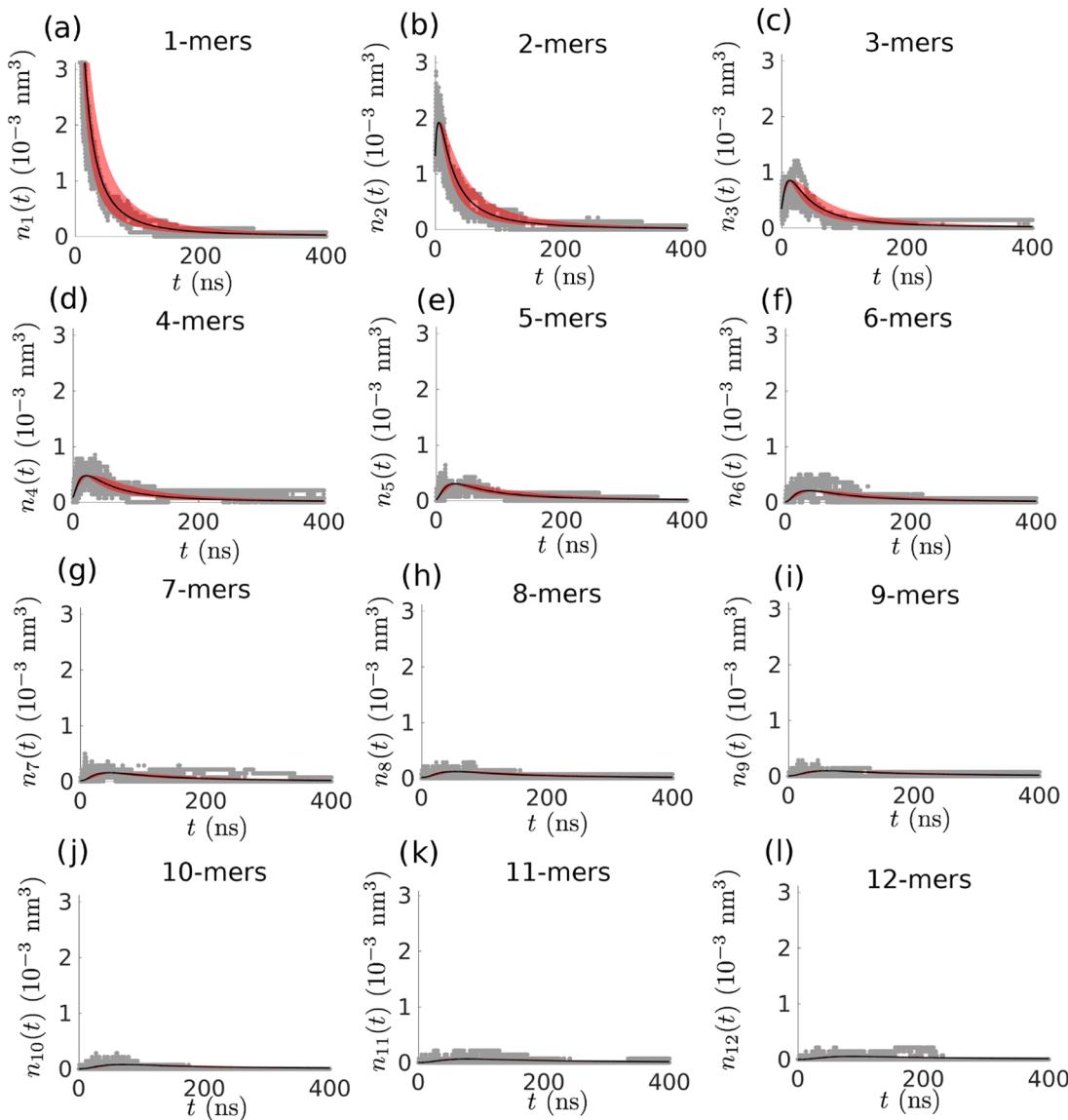


Figure 11. Comparison of time evolution of cluster number concentrations observed in simulation with those predicted by the fitted Smoluchowski coagulation equation for the aggregation of protonated (low-pH) peptides at 24 mM. (a–l) Number concentration time courses for monomers through 12-mers. Simulation data are displayed in gray for the five separate runs. The analytical predictions of the Smoluchowski coagulation model employing the best-fit value of $K = 9 \text{ ns}^{-1} \text{ nm}^3$ (Table 1) are presented as a dashed black line. Red shading illustrates the sensitivity of the model predictions to K over the range of $K \pm \sigma_K$ where σ_K is the standard deviation of K computed over the five independent production runs. Analogous plots for the four other concentrations studied are presented in Figures S6–S9.

size smaller than 20 in Figure 13 because, as evinced in Figure 14a–d, the scaling relation is not predicted to hold for small clusters and/or early times. In testing the scaling hypothesis for our simulation data, there is an inherent trade-off between analyzing small clusters, where we possess many observations, and analyzing large clusters, where the scaling relation is expected to be most valid. Because of a relative paucity of observational data for cluster sizes in excess of ~ 50 , we restrict our analysis in Figure 14 to clusters of a maximum size of 30. The fact that we observe approximate data collapse onto a master curve, $\phi(r/\mu_2(t))$, for sufficiently large clusters for which we possess sufficient statistics provides further support for the validity of the Smoluchowski coagulation model and self-similarity in the aggregation dynamics at multiple concentrations. There are two key implications of this result. First, the initial conditions do not strongly affect the long-time dynamics of the assembly. In particular, the cluster size distributions at

long times ($t \gg t_c$) do not depend on whether the initial system is composed of monodisperse monomers, as employed in our simulations, or a distribution of oligomers, as is a more realistic description of the high-pH aggregation state of the system (cf. Figure 4a). Second, the aggregation mechanism does not change as a function of concentration. Monomer concentration affects the rate of aggregation (cf. Figure 12) but not its underlying mechanism over the 5–44 mM concentration range studied in this work.

3.3. Analytical Predictions of the Smoluchowski Model. Assuming that the underlying mechanisms of the assembly remain unchanged for larger clusters, we may extrapolate our Smoluchowski coagulation equation fitted for the aggregation of hundreds of peptides over hundreds of nanoseconds to make predictions of the assembly at longer time and length scales. It is conceivable that additional physical concerns come into play for sufficiently large clusters that

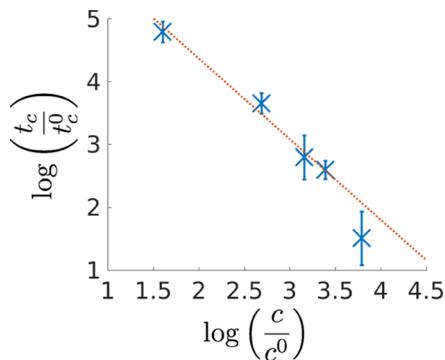


Figure 12. Log–log plot of coagulation time t_c as a function of initial monomer concentration. The dotted red line represents a least-squares fit of a power law form $t_c(c) = \zeta c^{-\gamma}$, where each data point is weighted in inverse proportion to its variance, from which we estimate $\gamma = 1.3$ (95% CI: 0.9, 1.7) and $\zeta = 1000$ (95% CI: 400, 3000). We render t_c and c dimensionless before taking logarithms by rescaling by the arbitrary reference time $t_c^0 = 1$ ns and concentration $c^0 = 1$ mM.

qualitatively change the structure and dynamics of the assembly; however, in the absence of data in very large cluster size regimes, we quantify the uncertainties in our extrapolation

on the basis of the uncertainty in the coagulation rate constant (K) estimated from the least-squares fit of the concentration-dependent characteristic coagulation time (Figure 12). In Figure 15, we present the analytical Smoluchowski predictions for the time evolution of clusters of size $r = 10^2$ (~ 10 nm), 10^3 (~ 100 nm), 10^4 (~ 1000 nm), 10^5 (~ 10000 nm), and 10^6 (~ 100000 nm) using the analytical solutions for $n_r(t)$ in eq 16 at a concentration $c = 0.86$ mM (0.1 mg/mL, $M_1 = 5 \times 10^{-4}$ nm $^{-3}$), previously employed in the directed microfluidic assembly of DFAG-OPV3-GAFD peptides,¹⁹ and a value of $K = 5$ ns $^{-1}$ nm 3 (95% CI: 3 ns $^{-1}$ nm 3 , 7 ns $^{-1}$ nm 3) estimated from the concentration-dependent extrapolation in Figure 12. Estimates of the approximate linear dimension of each cluster size were made from simple extrapolations of the observation that clusters of ~ 100 monomers in our simulations possess typical linear dimensions of ~ 10 nm. The confidence intervals in K were estimated by resampling the t_c values plotted in Figure 12 from Gaussians with variances matched to the standard deviations in the data and recomputing the least-squares fit from which the value of K at 0.86 mM was extrapolated over 100 rounds of this bootstrap resampling procedure.

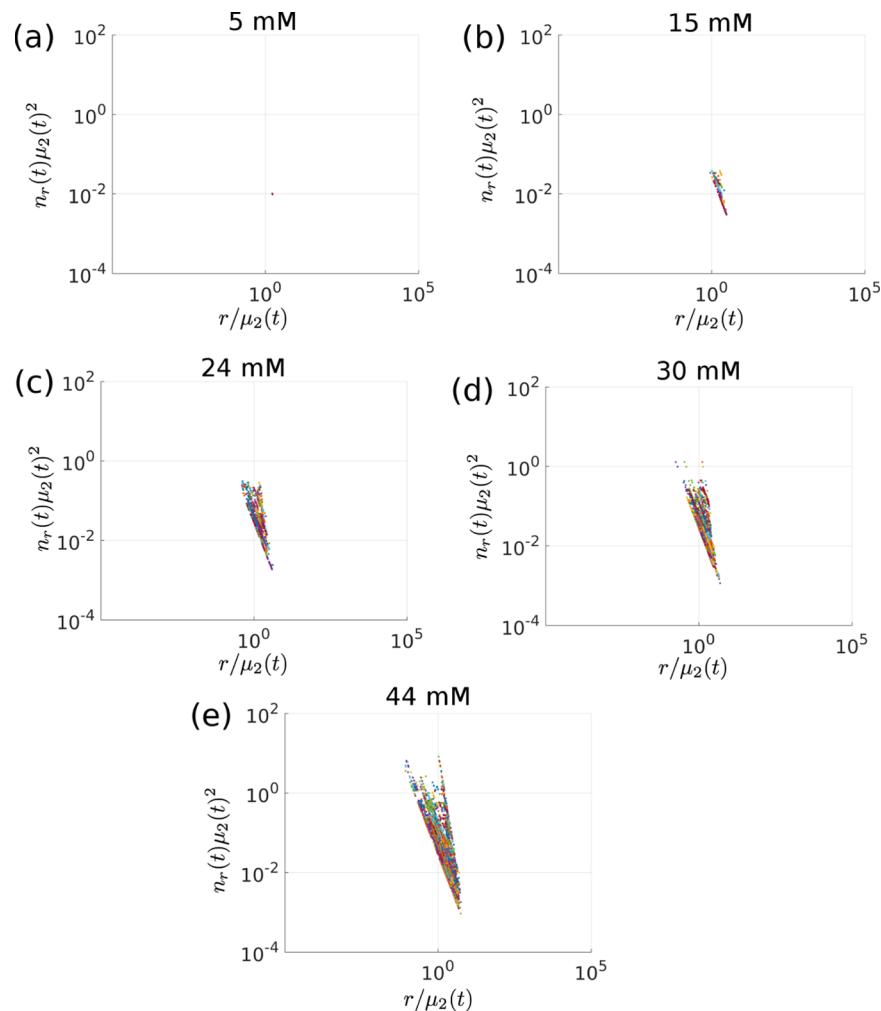


Figure 13. Plot of $n_r(t)\mu_2(t)^2$ against $r/\mu_2(t)$ at concentrations of (a) 5.0 mM, (b) 15 mM, (c) 24 mM, (d) 30 mM, and (e) 44 mM. Data are presented for clusters of size 20 and larger where the long-time scaling relation, eq 27, approximately holds. We observe approximate data collapse onto a master curve, $\phi(r/\mu_2(t))$. The visually apparent diagonal “streaking” occurs due to the discrete nature of cluster sizes.

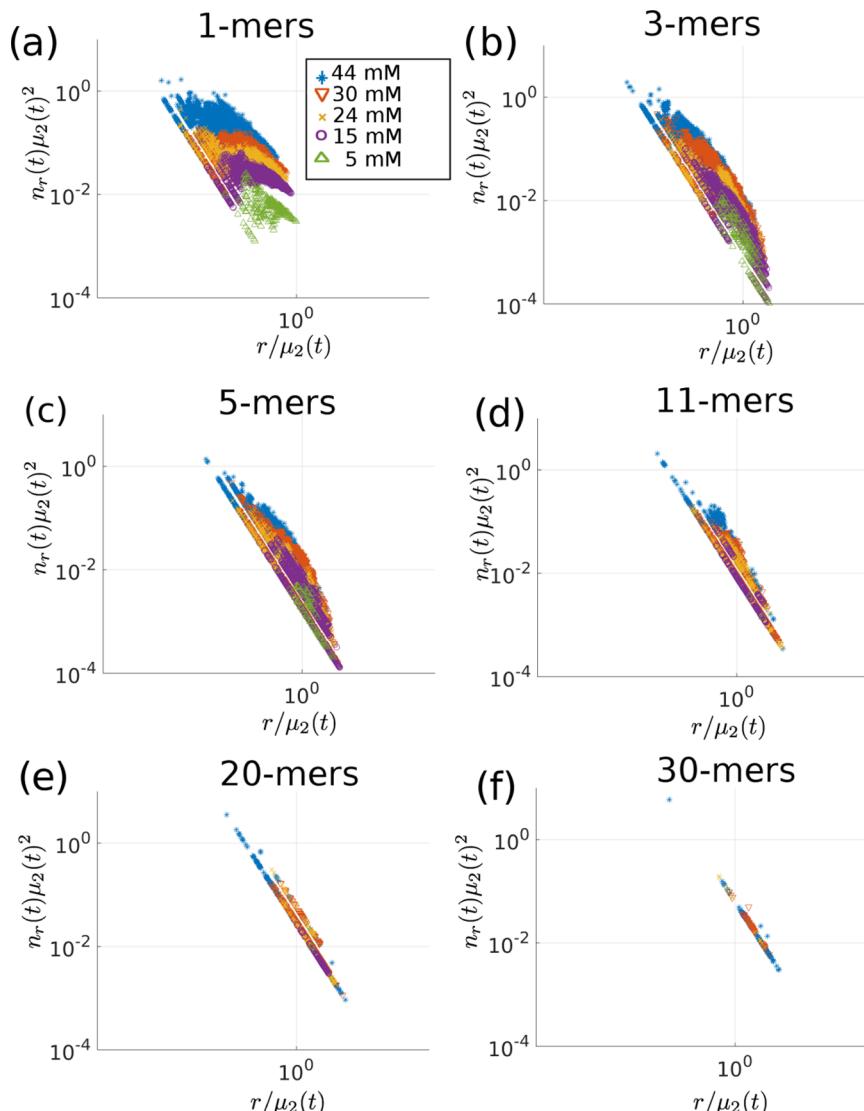


Figure 14. Plot of $n_r(t) \mu_2(t)^2$ against $r/\mu_2(t)$ observed in simulations at all concentrations studied for clusters of size (a) 1, (b) 3, (c) 5, (d) 11, (e) 20, and (f) 30. We observe approximate data collapse onto a master curve, $\phi(r/\mu_2(t))$, where the long-time scaling relation, eq 27, serves as a better approximation for larger clusters formed at later times. The visually apparent diagonal streaking occurs due to the discrete nature of cluster sizes.

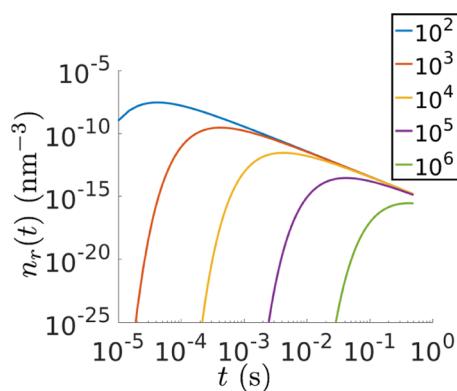


Figure 15. Analytical Smoluchowski predictions using eq 16 of the number concentrations, $n_r(t)$, of aggregates of sizes $r = 10^2$, $r = 10^3$, $r = 10^4$, $r = 10^5$, and $r = 10^6$ vs time t at a concentration $c = 0.86$ mM (0.1 mg/mL, $M_1 = 5 \times 10^{-4}$ nm $^{-3}$). The solutions converge at long times in line with the predictions of the scaling hypothesis.⁷³

The curves in Figure 15 represent theoretical predictions of the dynamical evolution of selected cluster sizes from our extrapolated Smoluchowski coagulation model, all of which collapse to the same curve, as expected from the scaling hypothesis that holds at long times and large clusters.⁷³ As it is extremely challenging to experimentally track the dynamics of particular cluster sizes, we cannot compare these predictions directly to experiment but can make experimental contact on a coarser level through the mass-averaged cluster size. In the directed assembly of aligned DFAG-OPV3-GAFD aggregates in a microfluidic flow cell at $c = 0.86$ mM (0.1 mg/mL), Marciel et al.¹⁹ reported the assembly and fluorescent detection of assemblies of diameter ~ 10 μ m on time scales of ~ 10 s. Using the same approximate conversion between cluster size and linear dimension as described above, these clusters are expected to contain $O(10^5)$ monomers. Rearranging the analytical expression for the mass-averaged cluster size in eq 18, we can estimate the time required for the coagulation process to produce a particular mass-averaged cluster size as

$$t(c) = \frac{(\mu_2 - \mu_2(0))}{KM_1} = \frac{t_c(c)}{2}(\mu_2 - \mu_2(0)) \quad (28)$$

where $t_c(c=0.86 \text{ mM}) = 800 \text{ ns}$ ($K(c = 0.86 \text{ mM}) = 5 \text{ ns}^{-1} \text{ nm}^3$) is extracted from the least-squares fit in Figure 12 and $\mu_2(0)$ is the initial mass-averaged cluster size. The experiments of Marciel et al.¹⁹ triggered assembly by combining an initially high-pH stream containing the peptides with a low-pH acid stream to drop the pH and initiate assembly. We approximately replicate this initial state by estimating $\mu_2(0) = (3.08 \pm 0.03)$ from the final 100 ns of the equilibrated deprotonated simulation to represent the high-pH aggregation state.

Solving eq 28 with $\mu_2 = 10^5$, $\mu_2(0) = 3.08$, and $t_c(c = 0.86 \text{ mM}) = 800 \text{ ns}$, we predict that the time required for the spontaneous assembly of aggregates of this size is 40 ms (95% CI: 10, 70 ms). Our estimate is 3 orders of magnitude slower than that reported by Marciel et al.,¹⁹ suggesting that there are slow dynamical processes occurring in the flow cell that are not captured within our Smoluchowski coagulation model. We hypothesize that these slow processes correspond to flow-induced ripening of the contact clusters through internal rearrangements to form larger and better ordered optical clusters with $\pi-\pi$ stacking of the OPV3 cores to form assemblies possessing fluorescent responses (cf. Figure 5). Contact clusters lacking this alignment are not expected to strongly fluoresce, suggesting that aggregation and fluorescence are not coextensive and that alternative experimental techniques would be required to determine contact cluster formation as a distinct process from optical cluster formation. Furthermore, the experimentally reported morphologies of the aligned aggregates under the extensional flow field employed in the microfluidic cell and under quiescent conditions are known to possess significant structural differences,¹⁹ indicating an important coupling of flow to the mechanisms and dynamics of assembly on large length scales. Together, these findings suggest a need for additional computational work to make better contact with the experimental work by employing electronic structure calculations to probe the different fluorescent properties of contact and optical clusters and more highly coarse-grained mesoscale models capable of probing the effect of flow on peptide assembly at longer time and length scales.

4. CONCLUSIONS

We have presented a computational study to probe the molecular details of the early-stage self-assembly of Asp-Phe-Ala-Gly-OPV3-Gly-Ala-Phe-Asp (DFAG-OPV3-GAFD) monomers containing an OPV3 (distyrylbenzene) core as a π -conjugated biocompatible oligopeptide that undergoes pH-triggered self-assembly into β -sheet-like nanostructures with experimentally observed optoelectronic functionality. We developed a coarse-grained molecular model for this system explicitly parameterized against all-atom calculations that recapitulates the configurational statistics of the intramolecular bonded interactions, intrapeptide contact maps, and PMFs for peptide collapse and dimerization observed in atomistic calculations. This protocol can be straightforwardly extended to develop coarse-grained models of other π -conjugated molecules with different peptide and core chemistries.

Using our computationally efficient coarse-grained model, we conducted molecular dynamics simulations of the assembly of hundreds of peptides over hundreds of nanoseconds. Morphological characterization shows that the initial stages of

the assembly proceed hierarchically. Monomers first form oligomers of ≤ 8 peptides with aligned aromatic cores (aligned clusters), which subsequently assemble into larger aggregates of $\lesssim 40$ peptides exhibiting strong associations between the π -conjugated cores (optical clusters). Optical clusters associate irreversibly via side-chain interactions to form large relatively disordered aggregates that we refer to as contact clusters. Contact clusters initially form into elongated ellipses with fractal dimension $D \approx 1.5$, which begin to display a branching behavior at peptide numbers of approximately 50–100. Internal rearrangements with these contact clusters produce increasingly large optical clusters but at a far lower rate than the formation of the contact clusters themselves. Experimental work has shown the importance of flow in ordering the self-assembled peptide aggregates on large length scales.^{17,19} Our molecular simulations suggest a role for flow in suppressing branching behavior and network formation to rotationally align the contact clusters into linear fibrillar bundles with improved $\pi-\pi$ stacking and optoelectronic properties.

Kinetically, we have shown the aggregation of contact clusters to be well described by a Smoluchowski coagulation model as a predictive phenomenological model of a nonequilibrium assembly.⁷³ Assuming a size-independent coagulation rate constant, we estimated this single fitting parameter from our molecular simulation data to establish analytical expressions for the time evolution of arbitrary cluster sizes and moments of the cluster size distribution. We demonstrated using the predicted scaling behavior of the Smoluchowski model that the aggregation dynamics observed in our simulations are self-similar, indicating that the long-time assembly behavior depends only weakly on the initial conditions and the assembly mechanism does not change with concentration over the 5–44 mM range studied in this work. Furthermore, we established a calibration curve to estimate the coagulation rate constant as a function of concentration, permitting the inexpensive analytical prediction of the microscopic cluster size evolution using the analytical predictions of the Smoluchowski model without the need to conduct any additional molecular simulations.

By revealing the molecular details of the hierarchical assembly mechanism over hundreds of nanoseconds and developing a phenomenological coagulation model of the nonequilibrium assembly kinetics, our work provides new fundamental understanding of peptide self-assembly and a route to rationally engineer π -conjugated peptide chemistry and environmental conditions to tune the morphology, kinetics, and optoelectronic properties of the emergent supramolecular assemblies. In future work, we will employ our modeling protocol to develop coarse-grained models of other oligopeptide chemistries containing different peptide sequences or π -conjugated cores (e.g., DXXX-II-XXXD, where X = {G, A, I, V, F} and II includes oligophenylvinylanes, oligothiophenes, and rylene diimides^{15,18,86}) and explicitly incorporate the effects of flow by conducting nonequilibrium molecular dynamics simulations implementing the Lees–Edwards⁸⁷ or Kraynik–Reinelt⁸⁸ boundary conditions.⁸⁹ Electronic structure calculations on representative aligned, optical, and contact clusters observed in this work will be informative in ascertaining the optoelectronic properties of the aggregates at each stage of the assembly hierarchy.^{90–93} We are also working to incorporate the Smoluchowski coagulation model fitted against our molecular simulation data into a multiscale model of peptide assembly within a microfluidic flow cell to enable explicit

comparisons with experimental measurements and provide a predictive model for peptide assembly as a function of peptide chemistry and the architecture and operating conditions of the microfluidic device.¹⁹

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jpcb.6b10165](https://doi.org/10.1021/acs.jpcb.6b10165).

Figures containing bonded distributions for reparameterized coarse-grained and atomistic systems, force-field parameters, and comparison of Smoluchowski predictions to simulation ([PDF](#))

Files b0–b8.txt and a0–a12.txt contain the tabulated force-field files and file topology.txt contains the DFAG coarse-grained topology ([ZIP](#))

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alf@illinois.edu. Phone: (217) 300-2354. Fax: (217) 333-2736.

ORCID

Andrew L. Ferguson: [0000-0002-8829-9726](https://orcid.org/0000-0002-8829-9726)

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award # DE-SC0011847. R.A.M. acknowledges a graduate fellowship from the Computational Science and Engineering Program at the University of Illinois. The authors thank Prof. Charles Schroeder, Prof. Jianjun Cheng, Dr. Bill Wilson, and Prof. Man Hoi Lee for fruitful discussions.

■ REFERENCES

- (1) Tuttle, T. Computational Approaches to Understanding the Self-assembly of Peptide-based Nanostructures. *Isr. J. Chem.* **2015**, *55*, 724–734.
- (2) Mehdizadeh, H.; Somo, S. I.; Bayrak, E. S.; Brey, E. M.; Cinar, A. Design of Polymer Scaffolds for Tissue Engineering Applications. *Ind. Eng. Chem. Res.* **2015**, *54*, 2317–2328.
- (3) Subramani, K.; Ahmed, W. *Emerging Nanotechnologies in Dentistry: Processes, Materials and Applications*; William Andrew: Waltham, 2011.
- (4) Hosseinkhani, H.; Hong, P.-D.; Yu, D.-S. Self-assembled Proteins and Peptides for Regenerative Medicine. *Chem. Rev.* **2013**, *113*, 4837–4861.
- (5) Palmer, L. C.; Newcomb, C. J.; Kaltz, S. R.; Spoerke, E. D.; Stupp, S. I. Biomimetic Systems for Hydroxyapatite Mineralization Inspired by Bone and Enamel. *Chem. Rev.* **2008**, *108*, 4754–4783.
- (6) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. Self-assembly and Mineralization of Peptide-Amphiphile Nanofibers. *Science* **2001**, *294*, 1684–1688.
- (7) Löwik, D. W.; Leunissen, E.; Van den Heuvel, M.; Hansen, M.; van Hest, J. C. Stimulus Responsive Peptide Based Materials. *Chem. Soc. Rev.* **2010**, *39*, 3394–3412.
- (8) Ulijn, R. V.; Smith, A. M. Designing Peptide Based Nanomaterials. *Chem. Soc. Rev.* **2008**, *37*, 664–675.
- (9) Bull, S. R.; Guler, M. O.; Bras, R. E.; Meade, T. J.; Stupp, S. I. Self-assembled Peptide Amphiphile Nanofibers Conjugated to MRI Contrast Agents. *Nano Lett.* **2005**, *5*, 1–4.
- (10) Zelzer, M.; Ulijn, R. V. Next-generation Peptide Nanomaterials: Molecular Networks, Interfaces and Supramolecular Functionality. *Chem. Soc. Rev.* **2010**, *39*, 3351–3357.
- (11) Guo, X.; Baumgarten, M.; Müllen, K. Designing π -Conjugated Polymers for Organic Electronics. *Prog. Polym. Sci.* **2013**, *38*, 1832–1908.
- (12) Bian, L.; Zhu, E.; Tang, J.; Tang, W.; Zhang, F. Recent Progress in the Design of Narrow Bandgap Conjugated Polymers for High-efficiency Organic Solar Cells. *Prog. Polym. Sci.* **2012**, *37*, 1292–1331.
- (13) Kim, S. H.; Parquette, J. R. A Model for the Controlled Assembly of Semiconductor Peptides. *Nanoscale* **2012**, *4*, 6940–6947.
- (14) Hoeben, F. J.; Jonkheijm, P.; Meijer, E.; Schenning, A. P. About Supramolecular Assemblies of π -Conjugated Systems. *Chem. Rev.* **2005**, *105*, 1491–1546.
- (15) Vadehra, G. S.; Wall, B. D.; Diegelmann, S. R.; Tovar, J. D. On-resin Dimerization Incorporates a Diverse Array of π -Conjugated Functionality within Aqueous Self-assembling Peptide Backbones. *Chem. Commun.* **2010**, *46*, 3947–3949.
- (16) Wall, B. D.; Zacca, A. E.; Sanders, A. M.; Wilson, W. L.; Ferguson, A. L.; Tovar, J. D. Supramolecular Polymorphism: Tunable Electronic Interactions within π -Conjugated Peptide Nanostructures Dictated by Primary Amino Acid Sequence. *Langmuir* **2014**, *30*, 5946–5956.
- (17) Wall, B. D.; Tovar, J. D. Synthesis and Characterization of π -Conjugated Peptide-based Supramolecular Materials. *Pure Appl. Chem.* **2012**, *84*, 1039–1045.
- (18) Wall, B. D.; Zhou, Y.; Mei, S.; Ardoña, H. A. M.; Ferguson, A. L.; Tovar, J. D. Variation of Formal Hydrogen-Bonding Networks within Electronically Delocalized π -Conjugated Oligopeptide Nanostructures. *Langmuir* **2014**, *30*, 11375–11385.
- (19) Marciel, A. B.; Tanyeri, M.; Wall, B. D.; Tovar, J. D.; Schroeder, C. M.; Wilson, W. L. Fluidic-Directed Assembly of Aligned Oligopeptides with π -Conjugated Cores. *Adv. Mater.* **2013**, *25*, 6398–6404.
- (20) Thurston, B. A.; Tovar, J. D.; Ferguson, A. L. Thermodynamics, Morphology, and Kinetics of Early-stage Self-assembly of π -Conjugated Oligopeptides. *Mol. Simul.* **2016**, *42*, 955–975.
- (21) Fitzpatrick, A. W.; et al. Atomic Structure and Hierarchical Assembly of a Cross- β Amyloid Fibril. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5468–5473.
- (22) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (23) de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (24) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. Coarse-grained Molecular Dynamics Simulations of Membrane Proteins and Peptides. *J. Struct. Biol.* **2007**, *157*, 593–605.
- (25) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. Coarse Grain Models and the Computer Simulation of Soft Materials. *J. Phys.: Condens. Matter* **2004**, *16*, R481.
- (26) Noid, W. Perspective: Coarse-grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139*, No. 090901.
- (27) Derreumaux, P.; Mousseau, N. Coarse-grained Protein Molecular Dynamics Simulations. *J. Chem. Phys.* **2007**, *126*, No. 025101.
- (28) Tozzini, V. Coarse-grained Models for Proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
- (29) Wang, Y.; Rader, A.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147*, 302–314.
- (30) Arkhipov, A.; Freddolino, P. L.; Imada, K.; Namba, K.; Schulten, K. Coarse-grained Molecular Dynamics Simulations of a Rotating Bacterial Flagellum. *Biophys. J.* **2006**, *91*, 4589–4597.

- (31) Wheeler, R. A.; Spellmeyer, D. C. *Annual Reports in Computational Chemistry*; Elsevier: Amsterdam, The Netherlands, 2010; Vol. 7.
- (32) Gautieri, A.; Russo, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J. Coarse-grained Model of Collagen Molecules Using an Extended MARTINI Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1210–1218.
- (33) Pannuzzo, M.; De Jong, D. H.; Raudino, A.; Marrink, S. J. Simulation of Polyethylene Glycol and Calcium-mediated Membrane Fusion. *J. Chem. Phys.* **2014**, *140*, No. 124905.
- (34) López, C. A.; de Vries, A. H.; Marrink, S. J. Computational Microscopy of Cyclodextrin Mediated Cholesterol Extraction from Lipid Model Membranes. *Sci. Rep.* **2013**, *3*, No. 2071.
- (35) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Probing the Self-assembly Mechanism of Diphenylalanine-based Peptide Nanovesicles and Nanotubes. *ACS Nano* **2012**, *6*, 3907–3918.
- (36) Seo, M.; Rauscher, S.; Pomès, R.; Tieleman, D. P. Improving Internal Peptide Dynamics in the Coarse-grained MARTINI Model: Toward Large-scale Simulations of Amyloid and Elastin-like Peptides. *J. Chem. Theory Comput.* **2012**, *8*, 1774–1785.
- (37) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.
- (38) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (39) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: a Tool for High-throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355–1363.
- (40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (41) MacKerell, A. D.; et al. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (42) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (43) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (44) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press: Bristol, UK, 2010.
- (45) 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*, 1463–1472.
- (46) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (47) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: Oxford, UK, 1989; p 21.
- (48) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (50) Periole, X.; Marrink, S.-J. *Biomolecular Simulations: Methods and Protocols*; Springer: New York City, NY, 2013; pp 533–565.
- (51) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (52) Sippl, M. J. Calculation of Conformational Ensembles from Potentials of Mean Force: an Approach to the Knowledge-based Prediction of Local Structures in Globular Proteins. *J. Mol. Biol.* **1990**, *213*, 859–883.
- (53) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-graining Applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.
- (54) MATLAB, version 9.0 (R2016a); The MathWorks Inc.: Natick, MA, 2016.
- (55) 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.
- (56) Pall, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Solving Software Challenges for Exascale*; Springer International Publishing: Cham, Switzerland, 2014; pp 3–27.
- (57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
- (58) Neumann, R. M. Entropic Approach to Brownian Movement. *Am. J. Phys.* **1980**, *48*, 354–357.
- (59) Ardoña, H. A. M.; Tovar, J. D. Energy Transfer Within Responsive pi-Conjugated Coassembled Peptide-based Nanostructures in Aqueous Environments. *Chem. Sci.* **2015**, *6*, 1474–1484.
- (60) Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F.; Kremer, K. Multiscale Modeling of Soft Matter: Scaling of Dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10412–10420.
- (61) Groot, R. D. Electrostatic Interactions in Dissipative Particle Dynamics-Simulation of Polyelectrolytes and Anionic Surfactants. *J. Chem. Phys.* **2003**, *118*, 11265–11277.
- (62) Groot, R. D.; Rabone, K. Mesoscopic Simulation of Cell Membrane Damage, Morphology Change and Rupture by Nonionic Surfactants. *Biophys. J.* **2001**, *81*, 725–736.
- (63) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (64) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Elsevier: San Diego, CA, 2002.
- (65) Wang, J.; Ferguson, A. L. Mesoscale Simulation of Asphaltene Aggregation. *J. Phys. Chem. B* **2016**, *120*, 8016–8035.
- (66) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
- (67) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (68) Grassberger, P.; Procaccia, I. Measuring the Strangeness of Strange Attractors. *Phys. D* **1983**, *9*, 189–208.
- (69) Theiler, J. Efficient Algorithm for Estimating the Correlation Dimension from a Set of Discrete Points. *Phys. Rev. A* **1987**, *36*, 4456.
- (70) Theodorou, D. N.; Suter, U. W. Shape of Unperturbed Linear Polymers: Polypropylene. *Macromolecules* **1985**, *18*, 1206–1214.
- (71) Smoluchowski, M. Drei Vortrage über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Z. Phys.* **1916**, *17*, 557–585.
- (72) Leyvraz, F. Scaling Theory and Exactly Solved Models in the Kinetics of Irreversible Aggregation. *Phys. Rep.* **2003**, *383*, 95–212.
- (73) Wattis, J. A. An Introduction to Mathematical Models of Coagulation-Fragmentation Processes: a Discrete Deterministic Mean-field Approach. *Phys. D* **2006**, *222*, 1–20.
- (74) Lee, M. H. A Survey of Numerical Solutions to the Coagulation Equation. *J. Phys. A: Math. Gen.* **2001**, *34*, 10219.
- (75) Lee, M. H. On the Validity of the Coagulation Equation and the Nature of Runaway Growth. *Icarus* **2000**, *143*, 74–86.
- (76) Olivier, B. J.; Sorensen, C.; Taylor, T. W. Scaling Dynamics of Aerosol Coagulation. *Phys. Rev. A* **1992**, *45*, S614.
- (77) Modler, A. J.; Gast, K.; Lutsch, G.; Damaschun, G. Assembly of Amyloid Protofibrils via Critical Oligomers—a Novel Pathway of Amyloid Formation. *J. Mol. Biol.* **2003**, *325*, 135–148.
- (78) Gillam, J. E.; MacPhee, C. Modelling Amyloid Fibril Formation Kinetics: Mechanisms of Nucleation and Growth. *J. Phys.: Condens. Matter* **2013**, *25*, No. 373101.
- (79) Treat, R. P. An Exact Solution of the Discrete Smoluchowski Equation and its Correspondence to the Solution of the Continuous Equation. *J. Phys. A: Math. Gen.* **1990**, *23*, 3003.
- (80) Leyvraz, F.; Tschudi, H. Singularities in the Kinetics of Coagulation Processes. *J. Phys. A: Math. Gen.* **1981**, *14*, 3389.
- (81) Smoluchowski, M. Versuch Einer Mathematischen Theorie der Koagulationskinetik Kolloider Lösungen. *Pisma Mariana Smoluchowskiego* **1927**, *2*, 595–639.

- (82) Lenaz, G. In *Coenzyme Q: Molecular Mechanisms in Health and Disease*; CRC Press: New York City, NY, 2000; Chapter 1, pp 5–28.
- (83) Chatterjee, S.; Hadi, A. S. *Regression Analysis by Example*; John Wiley & Sons: Hoboken, NJ, 2015.
- (84) Olivier, B. J.; Sorensen, C. Variable Aggregation Rates in Colloidal Gold: Kernel Homogeneity Dependence on Aggregant Concentration. *Phys. Rev. A* **1990**, *41*, 2093.
- (85) Van Dongen, P. G.; Ernst, M. Dynamic Scaling in the Kinetics of Clustering. *Phys. Rev. Lett.* **1985**, *54*, 1396.
- (86) Diegelmann, S. R.; Gorham, J. M.; Tovar, J. D. One-dimensional Optoelectronic Nanostructures Derived from the Aqueous Self-assembly of π -Conjugated Oligopeptides. *J. Am. Chem. Soc.* **2008**, *130*, 13840–13841.
- (87) Lees, A.; Edwards, S. The Computer Study of Transport Processes Under Extreme Conditions. *J. Phys. C: Solid State Phys.* **1972**, *5*, 1921.
- (88) Kraynik, A.; Reinelt, D. Extensional Motions of Spatially Periodic Lattices. *Int. J. Multiphase Flow* **1992**, *18*, 1045–1059.
- (89) Evans, D. J.; Morris, G. Nonlinear-response Theory for Steady Planar Couette Flow. *Phys. Rev. A* **1984**, *30*, 1528.
- (90) Rühle, V.; Lukyanov, A.; May, F.; Schrader, M.; Vehoff, T.; Kirkpatrick, J.; Baumeier, B.; Andrienko, D. Microscopic Simulations of Charge Transport in Disordered Organic Semiconductors. *J. Chem. Theory Comput.* **2011**, *7*, 3335–3345.
- (91) Cornil, J.; Beljonne, D.; Calbert, J.-P.; Brédas, J.-L. Interchain Interactions in Organic π -Conjugated Materials: Impact on Electronic Structure, Optical Response, and Charge Transport. *Adv. Mater.* **2001**, *13*, 1053–1067.
- (92) Caruso, A.; Siegler, M. A.; Tovar, J. D. Synthesis of Functionalizable Boron-Containing π -Electron Materials that Incorporate Formally Aromatic Fused Borepin Rings. *Angew. Chem.* **2010**, *122*, 4309–4313.
- (93) Caruso, A., Jr.; Tovar, J. D. Conjugated “B-Entacenes”: Polycyclic Aromatics Containing Two Borepin Rings. *Org. Lett.* **2011**, *13*, 3106–3109.