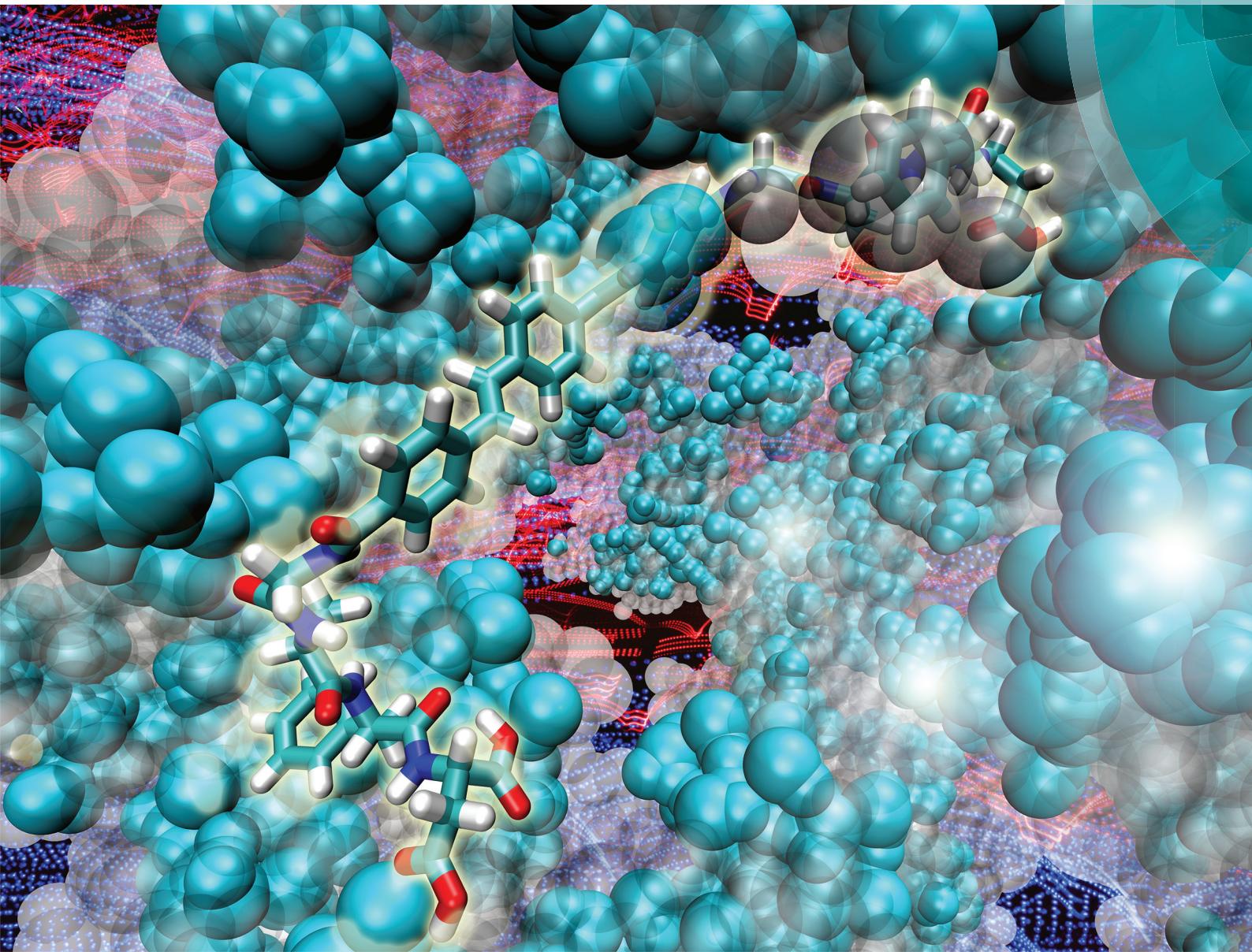


# Organic & Biomolecular Chemistry

rsc.li/obc



ISSN 1477-0520



ROYAL SOCIETY  
OF CHEMISTRY

PAPER

Rachael A. Mansbach and Andrew L. Ferguson  
Control of the hierarchical assembly of  $\pi$ -conjugated optoelectronic peptides by pH and flow



Cite this: *Org. Biomol. Chem.*, 2017,  
15, 5484

## Control of the hierarchical assembly of $\pi$ -conjugated optoelectronic peptides by pH and flow†

Rachael A. Mansbach <sup>a</sup> and Andrew L. Ferguson \*<sup>b,c</sup>

Self-assembled nanoaggregates of  $\pi$ -conjugated peptides possess optoelectronic properties due to electron delocalization over the conjugated peptide groups that make them attractive candidates for the fabrication of bioelectronic materials. We present a computational and theoretical study to resolve the microscopic effects of pH and flow on the non-equilibrium morphology and kinetics of early-stage assembly of an experimentally-realizable optoelectronic peptide that displays pH triggerable assembly. Employing coarse-grained molecular dynamics simulations, we probe the effects of pH on growth kinetics and aggregate morphology to show that control of the peptide protonation state by pH can be used to modulate the assembly rates, degree of molecular alignment, and resulting morphologies within the self-assembling nanoaggregates. We also quantify the time and length scales at which convective flows employed in directed assembly compete with microscopic diffusion to show that flow influences cluster alignment and assembly rate during early-stage assembly only at extremely high shear rates. This suggests that observed improvements in optoelectronic properties at experimentally-accessible shear rates are due to the alignment of large aggregates of hundreds of monomers on time scales in excess of hundreds of nanoseconds. Our work provides new fundamental understanding of the effects of pH and flow to control the morphology and kinetics of early-stage assembly of  $\pi$ -conjugated peptides and lays the groundwork for the rational manipulation of environmental conditions to direct assembly and the attendant emergent optoelectronic properties.

Received 13th April 2017,  
Accepted 13th May 2017

DOI: 10.1039/c7ob00923b  
[rsc.li/obc](http://rsc.li/obc)

## 1 Introduction

Well-ordered peptide and polymer aggregates with aromatic functional groups are desirable targets for the design of tunable electronic biomaterials such as organic photovoltaics, light-emitting diodes, field effect transistors, and biocompatible sensors due to the optoelectronic properties they evince in ordered nanoaggregate assemblies due to intermolecular overlaps between aromatic  $\pi$  orbitals,<sup>1–4</sup> enhanced proton transfer along hydrogen bonds,<sup>5,6</sup> and/or combinations of more complicated structural and electronic effects.<sup>7</sup> Biomimetic self-assembling peptide aggregates are easily synthesizable and

their supramolecular morphology can be tuned by the assembly environment, such as the presence or absence of external flow, the pH or rate of change of pH, the presence or absence of metal ions, and exposure to light, providing a route to engineer the attendant optoelectronic properties.<sup>8–17</sup> Molecular-level understanding of the impact of the environment upon assembly is a prerequisite to inform the rational design of self-assembling nanoaggregates for specific applications under particular conditions.

The DXXX- $\Pi$ -XXXD family is a class of self-assembling peptides with aromatic cores and tunable optoelectronic properties that exhibits pH-triggerable assembly.<sup>11,12,18–21</sup> The DXXX refers to the amino acid sequence of the flanking tetrapeptide wings containing distal Asp residues that provide the pH trigger for assembly, and the  $\Pi$  to the identity of the  $\pi$ -conjugated insert serving as the molecular core. At high pH ( $\text{pH} > 7$ ), the two carboxyl groups in each Asp residue are deprotonated, endowing each molecule with a net charge of ( $-4$ ) that restricts aggregation to small oligomers and prohibits large-scale assembly due to repulsive electrostatic interactions. At low pH ( $\text{pH} < 1$ ), the Asp residues protonate to eliminate this electrostatic repulsion, and the peptides self-assemble

<sup>a</sup>Department of Physics, University of Illinois at Urbana-Champaign, 1110 W Green St, Urbana, IL 61801, USA

<sup>b</sup>Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, 1304 W Green Street, Urbana, IL 61801, USA

<sup>c</sup>Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA

E-mail: [alf@illinois.edu](mailto:alf@illinois.edu); Fax: +1 (217) 333-2736; Tel: +1 (217) 300-2354

† Electronic supplementary information (ESI) available: Plots of time-averaged slope of mass-averaged cluster size, fits for Smoluchowski analysis, and movies of five selected independent simulations. See DOI: 10.1039/C7OB00923B

into large pseudo-1D aggregates. Experimentally, it has been shown that environmental factors, such as the presence of external flow and the rapidity at which the pH is dropped affect the morphology and optoelectronic properties of the terminal peptide aggregates.<sup>11,13</sup> Several computational and theoretical studies of the DXXX family have been performed to quantify the equilibrium morphologies of preformed  $\beta$ -sheet-like aggregates,<sup>20</sup> the free energetics of dimer formation,<sup>19</sup> and quantification of the kinetics of early-stage assembly using a Markov state model.<sup>21</sup>

Recently, we developed a bottom-up coarse-grained model for the symmetric DFAG-OPV3-GAFD peptide explicitly parameterized against all-atom simulation data<sup>12</sup> (Fig. 1). By performing simulations of hundreds of monomers over hundreds of nanoseconds, we demonstrated that at low pH in quiescent aqueous solution the peptides undergo irreversible hierarchical aggregation. Monomers first aggregate into small oligomers of size  $\leq 8$  with well-aligned aromatic cores. These well-aligned small aggregates subsequently assemble through both core-to-core and promiscuous side chain interactions to form larger, more disordered aggregates. As we employed classical molecular dynamics simulations, we used proximity and alignment of the aromatic cores as a morphological proxy for optoelectronic functionality.

In this paper, we probe the effects of pH and non-equilibrium flow on the self-assembly of DFAG-OPV3-GAFD peptides using our coarse-grained molecular simulation model.<sup>12</sup> We investigate the effects of pH by simulating systems of DFAG-OPV3-GAFD containing different fractions of charged monomers to demonstrate that increasing pH simultaneously slows the growth rate and improves the alignment of the aggregates formed in early-stage assembly. We show that the growth rate is well modeled by a homogeneous Smoluchowski coagu-

lation model in which the effect of the pH can be treated as an approximately linear decrease in the growth exponent. We also present a simple phenomenological model and scaling argument to understand the effects of pH on the growth rate due to electrostatic shielding of the self-assembling aggregates by the charged peptide termini. We investigate the effects of flow by simulating assembly under different rates of shear deformation. We find that flow accelerates the assembly rate only at extremely high shear rates, and that in experimentally-realized shear flows within microfluidic devices the flow is only predicted to influence the morphology of clusters containing in excess of hundreds of monomers at time scales in excess of hundreds of nanoseconds. We understand the influence of flow as a function of time and length scale by performing a dimensionless analysis of the flow behavior using a Pipkin-type analysis familiar from rheology.<sup>22,23</sup>

The outline of this article is as follows. In section 2, we describe our simulation methods. In section 3, we present our results for the influence of pH and flow upon assembly. In section 4, we present our conclusions and directions for future work.

## 2 Methods

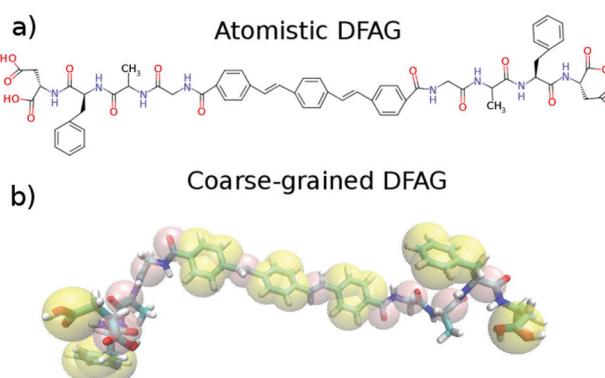
### 2.1 Coarse-grained simulations at different pH

Assembly in the DFAG-OPV3-GAFD system is triggered when the aspartic acid residues on the ends of each molecule become protonated, lowering the electrostatic repulsion between monomers as the net charge per monomer changes from ( $-4$ ) to  $0$ . We represent different pH levels by simulating systems with different fractions of charged monomers. This approach is computationally efficient since each molecule is assigned an immutable protonation state of either ( $-4$ ) or  $0$ , thereby neglecting both the possibility of intermediate charged states and the dynamic equilibrium of any particular molecule between a distribution of protonation states that may be a function of its local environment. We appreciate that a more sophisticated simulation approach would employ constant-pH ensemble simulations.<sup>26–30</sup> Although there have been some initial efforts to generalize constant pH simulations to the coarse-grained Martini model,<sup>31</sup> the high computational overhead associated with such calculations would prohibit access to the time and length scales necessary to observe early-stage assembly.

We define the relative fractions of protonated and deprotonated monomers as a function of pH using the Henderson-Hasselbalch equation,<sup>32</sup>

$$\text{pH} = \text{p}K_{\text{a}} + \log_{10} \left( \frac{[\text{A}^-]}{[\text{HA}]} \right), \quad (1)$$

where  $\text{p}K_{\text{a}} = 3.9$  pertains to the terminal aspartic acid residues,<sup>33</sup> and  $[\text{A}^-]/[\text{HA}]$  is the ratio of deprotonated to protonated monomers. We list in Table 1 the fraction of charged (deprotonated) monomers in the system as a function of the eight pH values considered in this work. The expected fraction



**Fig. 1** Chemical structure and coarse-grained model of the symmetric DFAG-OPV3-GAFD peptide. At high pH (pH  $> 7$ ), the distal aspartic acid residues are deprotonated and the molecule carries a ( $-4$ ) charge. At low pH (pH  $< 1$ ), the Asp residues protonate and the molecules become electrostatically neutral. (a) Chemical structure drawn using the Marvin software package, Marvin 15.12.7.0 (ChemAxon, 2015) (<http://www.chemaxon.com>), (b) mapping from atomistic structure to coarse-grained Martini beads<sup>24</sup> as shown by VMD.<sup>25</sup> The full description of the force field used and the force field files can be found in the ESI of ref. 12.

**Table 1** Correspondence between pH and the number of deprotonated (charged) monomers computed by the Henderson–Hasselbach relation (eqn (1)). Our simulations comprise a total of 378 monomers. The ionizable Asp termini have a  $pK_a = 3.9$  such that the system is fully protonated at  $\text{pH} < 1$  ( $\lesssim 0.1\%$  deprotonated) and fully deprotonated at  $\text{pH} > 7$  ( $\gtrsim 99.9\%$  deprotonated)

pH	% deprotonated monomers	No. deprotonated monomers
<1.0	0.00%	0
3.43	25.13%	95
3.68	37.57%	142
3.90	50.00%	189
4.12	62.43%	236
4.37	74.87%	283
4.75	87.57%	331
>7.0	100.00%	378

of deprotonated monomers is  $\sim 0.1\%$  at  $\text{pH} = 1$  and  $\sim 99.9\%$  at  $\text{pH} = 7$ . Our simulations comprise 378 total monomers such that the system is expected to be fully protonated at  $\text{pH} < 1$  (*i.e.*, fewer than 0.5 deprotonated monomers) and fully deprotonated at  $\text{pH} > 7$  (*i.e.*, fewer than 0.5 protonated monomers).

Simulations of coarse-grained DFAG-OPV3-GAFD systems at eight different pH values were conducted in the Gromacs 4.6 suite.<sup>34</sup> The monomers were represented by the modified Martini model detailed in ref. 12. Briefly, the coarse-grained model of DFAG-OPV3-GAFD (see Fig. 1b) consists of a set of beads as defined in the Martini force field,<sup>24,35</sup> wherein approximately two to four heavy atoms are mapped to a single bead. Intramolecular forces are taken from the Martini 2.2p force field.<sup>36</sup> Intermolecular forces are defined *via* Boltzmann Inversion performed on distributions taken from a 340 ns atomistic simulation 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).<sup>12,37,38</sup> Water was modeled as Martini polarizable water.<sup>39</sup> The complete force field and topology files are available as part of the ESI in ref. 12.

Peptides were prepared in two protonation states: a protonated state that was electrostatically neutral and a deprotonated state with a formal charge of  $(-4)$ , where the only difference between the two states is in the beads representing the terminal aspartic acid residues. In the protonated state, each two-bead terminus was represented by the Martini semi-attractive P3 and P5 beads, whereas in the deprotonated state, both terminal beads were represented by the Martini (negatively) charged Qa beads.

Systems of 0 deprotonated and 378 protonated ( $0.00\%$  charged molecules, corresponding to  $\text{pH} < 1$ ), 95 deprotonated and 283 protonated ( $25.13\%$  charged molecules, corresponding to  $\text{pH} \sim 3.43$ ), 142 deprotonated and 236 protonated ( $37.57\%$  charged molecules, corresponding to  $\text{pH} \sim 3.68$ ), 189 deprotonated and 189 protonated ( $50.00\%$  charged molecules, corresponding to  $\text{pH} \sim 3.90$ ), 236 deprotonated and 142 protonated ( $62.43\%$  charged molecules, corresponding to  $\text{pH} \sim 4.12$ ), 283 deprotonated and 95 protonated molecules ( $74.87\%$  charged molecules, corresponding to  $\text{pH} \sim 4.37$ ), 331

deprotonated and 47 protonated molecules ( $87.57\%$  charged molecules, corresponding to  $\text{pH} \sim 4.75$ ), and 378 deprotonated and 0 protonated molecules ( $100.00\%$  charged molecules, corresponding to  $\text{pH} > 7$ ) were prepared (Table 1). In each case, the DFAG-OPV3-GAFD molecules were inserted at random locations into a simulation box of size  $23.5 \times 23.5 \times 23.5 \text{ nm}^3$  and solvated in Martini polarizable water<sup>39</sup> at an effective density of  $\rho \approx 1 \text{ g cm}^{-3}$  to average initial number concentrations of  $\sim 50 \text{ mM}$ . Charge neutrality was maintained by randomly inserting a number of counter ions equal to the total negative charge in the system. Counter ions were represented by Martini Qd beads each carrying a  $+1$  charge.

Performing steepest descent energy minimization reduced forces to below  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  and removed high energy overlaps in the initial conformations of the solvated peptide systems. Simulations were conducted employing a velocity rescaling thermostat<sup>40</sup> and Parrinello–Rahman barostat<sup>41</sup> within the NPT ensemble at 298 K and 1 bar. Initial atom velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K. A leap-frog algorithm with a 5 fs time step was employed to numerically integrate the equations of motion.<sup>42</sup> The cut-off of the Lennard–Jones interactions was set to 1.1 nm, after which the interactions were smoothly shifted to zero. The reaction field method with  $\epsilon_{\text{rf}} = \infty$  and  $\epsilon_{\text{r}} = 2.5$  was used to model electrostatic interactions, as recommended for use with the Martini force-field and Martini polarizable water.<sup>39</sup> The values of the energy, density, temperature, and pressure attained stable values following an equilibration of 50 ps. Five independent production runs of 400 ns were conducted at each pH.

## 2.2 Coarse-grained simulations under shear flow

The initial configuration of the system was created by randomly inserting into a  $23.5 \times 23.5 \times 23.5 \text{ nm}^3$  box 200 protonated (uncharged) molecules distributed as 6 monomers, 21 dimers, 30 trimers, 13 tetramers, and 2 pentamers to approximately match the equilibrium distribution of the 100% deprotonated system.<sup>12</sup> This initial state is designed to represent the state of the system under a rapid drop in pH from  $\text{pH} > 7$  to  $\text{pH} < 1$  and mimic the conditions in a microfluidic flow cell in which a high-pH solution of peptides is flowed into the device and assembly triggered by mixing with an acid stream to drop the pH.<sup>11</sup>

Simulations were conducted employing the Gromacs 4.6 Suite<sup>34</sup> and the modified Martini model as described in the previous section, except that the intermolecular forces were taken from the Martini 2.2 force field and non-polarizable water was employed for higher efficiency.<sup>24</sup> Performing steepest descent energy minimization reduced forces to below  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  and removed high energy overlaps in the initial configurations of the solvated peptide systems. Simulations were conducted in the NPT ensemble at 298 K and 1 bar, employing a Nosé–Hoover thermostat<sup>43</sup> and Parrinello–Rahman barostat.<sup>41</sup> The Nosé–Hoover thermostat is favored over the velocity rescaling approach conventionally employed in coarse-grained Martini model calculations due to its

improved reproduction of hydrodynamic effects necessary to accurately model the effects of flow.<sup>44</sup> Initial bead velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K. A leap-frog algorithm with a 5 fs time step was employed to numerically integrate the equations of motion.<sup>42</sup> The cutoff of the Lennard-Jones interactions was set to 1.1 nm, at which distance the interactions were shifted smoothly to zero. Reaction-field electrostatics with a relative electrostatic screening of  $\epsilon_r = 15$  were used to model electrostatic forces. Energy, density, temperature, and pressure attained stable values after an equilibration period of 50 ps.

The application of shear flow to the system was modeled using the deform command in Gromacs to enforce Lees–Edwards boundary conditions.<sup>45</sup> Five independent production simulations of 400 ns were performed at each of six deformation rates of  $\dot{\delta} = 0 \text{ nm ps}^{-1}$ ,  $\dot{\delta} = 5 \times 10^{-7} \text{ nm ps}^{-1}$ ,  $\dot{\delta} = 5 \times 10^{-6} \text{ nm ps}^{-1}$ ,  $\dot{\delta} = 5 \times 10^{-5} \text{ nm ps}^{-1}$ ,  $\dot{\delta} = 5 \times 10^{-4} \text{ nm ps}^{-1}$ , and  $\dot{\delta} = 5 \times 10^{-3} \text{ nm ps}^{-1}$ , corresponding to shear rates of  $\dot{\gamma} = 0 \text{ s}^{-1}$ ,  $\dot{\gamma} = 2.1 \times 10^4 \text{ s}^{-1}$ ,  $\dot{\gamma} = 2.1 \times 10^5 \text{ s}^{-1}$ ,  $\dot{\gamma} = 2.1 \times 10^6 \text{ s}^{-1}$ ,  $\dot{\gamma} = 2.1 \times 10^7 \text{ s}^{-1}$ , and  $\dot{\gamma} = 2.1 \times 10^8 \text{ s}^{-1}$ , where  $\dot{\gamma} = \dot{\delta}/L$  and  $L = 23.5 \text{ nm}$  is the characteristic box length.<sup>46</sup> Employing solely Lees Edwards boundary conditions to simulate flow means that the velocity profile over the system does not change instantaneously, but rather propagates at roughly the speed of sound. Adopting a speed of sound of 1494 m s<sup>-1</sup> corresponding to that in water at 298 K,<sup>47</sup> it will take about 20 ps for the flow to propagate through a box with side length of 30 nm, and the short time delay associated with this transient is therefore of negligible concern over the course of our 400 ns production runs.

**2.2.1 Shear flows are in the laminar regime.** Reynolds number ( $Re$ ) calculations for the experimental assembly of these oligopeptides in microfluidic devices indicates that assembly proceeds in the laminar regime,<sup>11</sup> motivating our decision to model flow as simple shear deformation. We now verify that our simulations are approximately in the laminar flow regime at all shear rates considered. The Reynolds number is defined as,

$$Re \equiv \frac{\rho UL}{\mu}, \quad (2)$$

where  $\rho$  is the density,  $U$  is the average velocity of the fluid,  $L$  is a characteristic length scale, and  $\mu$  is the solvent viscosity. We adopt  $\mu = 7 \times 10^{-4} \text{ Pa s}$ , the viscosity of Martini water,<sup>48</sup>  $\rho = 1 \text{ g cm}^{-3}$ , the density of water under standard temperature and pressure, and  $L = 23.5 \text{ nm}$ , the box side length. The average fluid velocity under the fully-developed shear flow profile is computed as,<sup>46</sup>

$$U = \langle v \rangle = \dot{\gamma} \frac{L}{2}, \quad (3)$$

from which we estimate a maximum fluid velocity of  $U = 2.5 \times 10^{-3} \text{ nm ps}^{-1}$ . The maximum expected Reynolds numbers is then  $Re = 8 \times 10^{-2} \ll O(1)$ , putting us squarely in the laminar flow regime. We observe that the (non-zero) average velocities  $U$  considered in this work range from 0.025 cm s<sup>-1</sup> to 250 cm s<sup>-1</sup>. Experimental average velocities are approxi-

mately 0.3 cm s<sup>-1</sup>, corresponding to shear rates of  $\dot{\gamma} \approx 2.5 \times 10^5 \text{ s}^{-1}$ .<sup>11</sup>

### 2.3 Coarse-grained simulations of isolated aggregates

In our dimensionless analysis of shear flow we require an estimate of the rotational diffusivity  $D_r$  as a function of cluster size (see section 3.5). We empirically estimate this by harvesting 70 contact clusters of various sizes ranging from 6 to 330 monomers from the final time step of the 400 ns simulations of quiescent protonated (pH < 1.0) self-assembly described in ref. 12. We then placed each cluster in its own box of water to permit us to observe the rotational dynamics of isolated clusters. For efficiency, the box size in each case was chosen to be the smallest such that the clusters would not self-interact across the periodic boundaries. The box size employed for each cluster size is given in Table 2.

Calculations were carried out in Gromacs 4.6,<sup>34</sup> using the modified Martini coarse-grained model described above, with intermolecular interactions characterized by the Martini 2.2 force field and water represented by Martini non-polarizable water.<sup>24</sup> Simulations were conducted in the NPT ensemble at 298 K and 1 bar, employing a velocity rescaling thermostat,<sup>40</sup> a Parrinello–Rahman barostat,<sup>41</sup> and three-dimensional periodic boundary conditions. Initial bead velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K, and high energy overlaps in the initial configurations eliminated by performing steepest descent energy minimization to remove forces exceeding 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup>. Equations of motion were numerically integrated using a leap-frog algorithm<sup>42</sup> with a 5 fs time step, and constrained bond lengths were fixed using the LINCS algorithm.<sup>49</sup> Lennard-Jones interactions were shifted smoothly to zero at 1.1 nm. Reaction field electrostatics with a relative electrostatic screening of  $\epsilon_r = 15$  were used to model electrostatic forces. Temperature, pressure, and density reached stable values following equilibration simulations of 100 ps. A 1 ns production run was then conducted for each cluster.

The rotational diffusivity of each cluster was estimated from an Einstein relation describing the mean-squared angular displacement of the cluster orientation  $\theta$  as a function of time  $t$ ,<sup>50,51</sup>

$$\langle (\theta(t) - \theta_0)^2 \rangle = 2dD_r t, \quad (4)$$

**Table 2** Length of box side  $L$  for cubic triclinic box used to simulate isolated clusters comprising different numbers of monomers

Cluster size (no. monomers)	$L$ (nm)
6	7
11	8
16	9
20	10
30–39	12
40–59	15
60–74	18
75–179	31
180–330	33

where  $\theta_0$  is the initial orientation,  $d = 2$  is the appropriate dimensionality for a body free to rotate in three dimensions, and  $D_r$  is the rotational diffusivity. We define the orientation of a cluster from the unit vector  $\hat{\xi}_1$  coincident with the first principal eigenvector of the gyration tensor  $S$  of the cluster (see section 3.2.1 and eqn (11)), and we compute the angular displacement over a time interval  $\Delta t$  as,

$$\theta(t + \Delta t) - \theta(t) = \cos^{-1}(\hat{\xi}_1(t + \Delta t) \cdot \hat{\xi}_1(t)), \quad (5)$$

where the trajectory of  $\hat{\xi}_1$  is preprocessed to eliminate flips in the (arbitrary) sense of the principal eigenvector or spurious inversions of the first and second principal eigenvectors that can occur due to internal rearrangements within the roughly spherical small aggregates containing only six monomers.  $D_r$  was then estimated for each of the 70 clusters by making a linear least-squares fit to the linear regime of the mean-squared angular displacement curve and dividing the slope by  $2d = 4$ .<sup>52</sup>

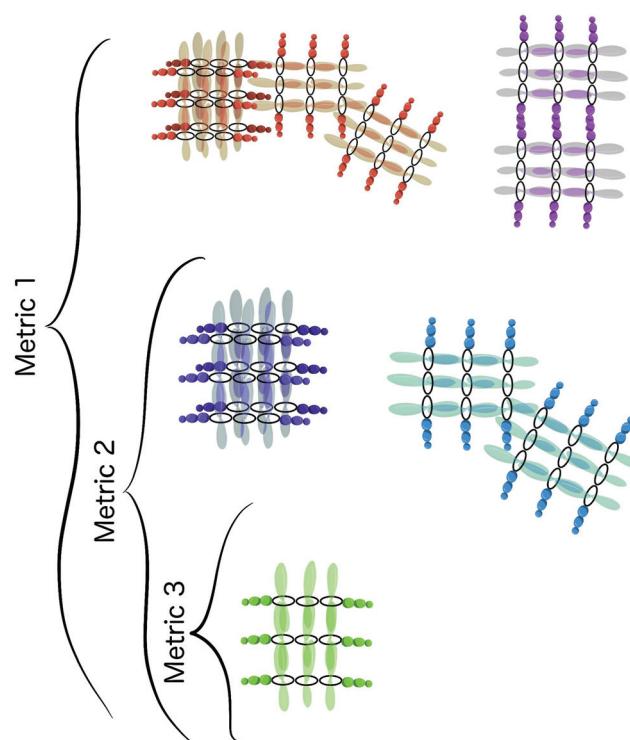
## 2.4 Definition of cluster types

We have previously shown DFAG-OPV3-GAFD peptides to undergo hierarchical aggregation at low pH in quiescent solution, and defined three different cluster types to characterize this hierarchical assembly (see Fig. 2), which we briefly describe below in order from most to least strict.<sup>12</sup> We note that these metrics are necessarily hierarchically nested, such that a pair of peptides identified as belonging to the same aligned cluster also belong to the same optical and contact cluster, and a pair identified as belonging to the same optical cluster necessarily also belong to the same contact cluster. Accordingly, the contact cluster defined for a particular ensemble of peptides is necessarily larger than or equal to the optical cluster, which is in turn constrained to be larger than or equal to the aligned cluster. We also observe the similarity of this hierarchical aggregation to the Yen–Mullins model for the aggregation of asphaltenes constituting the heaviest aromatic fraction in crude oil.<sup>53,54</sup>

**2.4.1 Aligned clusters.** Aligned clusters are defined by the distance metric,

$$R_{a,b}^{(3)} = \max \left[ \left( \max_{i \in (\text{core } a)} \min_{j \in (\text{core } b)} r_{ij} \right), \left( \max_{i' \in (\text{core } b)} \min_{j' \in (\text{core } a)} r_{i'j'} \right) \right], \quad (6)$$

where  $r_{ij}$  is the distance between the centers of mass of the  $i$ th aromatic ring in the core of molecule  $a$  and the  $j$ th aromatic ring in the core of molecule  $b$ . When the distance between molecules  $a$  and  $b$  is such that  $R_{a,b}^{(3)} < R_{\text{cut}}^{(3)} = 0.7$  nm,  $a$  and  $b$  are in the same aligned cluster. It should be noted that the maximin operation does not commute, hence why it is taken twice. As previously observed, this distance is analogous to the diameter of a graph.<sup>54</sup> Aligned clusters, as the name suggests, are aggregates in which the aromatic cores of each constituent monomer are mutually well aligned and expected to possess desirable optoelectronic properties due to the intimate intermolecular interactions between the aromatic cores.<sup>11,12,18–21</sup>



**Fig. 2** Schematic illustration of hierarchy of different cluster types. Metric 3 defines aligned clusters, which have aligned and strongly-interacting aromatic cores. Metric 2 defines optical clusters, which contain aligned clusters and clusters with more promiscuous core–core interactions. Metric 1 defines contact clusters, which contain the other two cluster types as well as clusters that interact through promiscuous core–side chain and side chain–side chain interactions. In this figure, black circles represent aromatic rings, shaded ellipsoids represent lobes of p orbitals above and below the plane of the aromatic rings, circular beads represent side chains, and colors distinguish the distinct clusters of each type at each level. This schematic presents an idealized representation of the aggregation behavior, and we report a variety of structural characterizations to precisely quantify aggregate morphology.

Monomers within aligned clusters possess nearest neighbor distances of  $\Delta R^{(3)} = (0.61 \pm 0.05)$  nm, and we have verified that there is no statistically significant change in this distribution as a function of pH or shear rate (2-sample Kolmogorov–Smirnov test;  $\alpha = 0.05$ ).

**2.4.2 Optical clusters.** Optical clusters are defined by the distance metric,

$$R_{a,b}^{(2)} = \min_{i \in (\text{core } a)} \min_{j \in (\text{core } b)} r_{ij}, \quad (7)$$

where  $r_{ij}$  is the distance between the centers of mass of any ring  $i$  in the core of molecule  $a$  and any ring  $j$  in the core of molecule  $b$ . When the distance between molecules  $a$  and  $b$  is such that  $R_{a,b}^{(2)} < R_{\text{cut}}^{(2)} = 0.7$  nm,  $a$  and  $b$  are in the same optical cluster. Therefore, optical clusters are defined as any cluster in which the cores of the constituent monomers are proximate, but not necessarily well aligned, as is required to be considered an aligned cluster. This cluster type contains both aligned clusters and clusters with more promiscuous

core–core interactions. Nevertheless, since all such clusters have core–core interactions of one kind or another, they are still candidates for desirable optoelectronic properties.

**2.4.3 Contact clusters.** Contact clusters are defined by the distance metric,

$$R_{a,b}^{(1)} = \min_{i \in a} \min_{j \in b} r_{ij}, \quad (8)$$

where  $r_{ij}$  is the distance between two beads  $i$  and  $j$  in monomers  $a$  and  $b$ , respectively. Monomers for which  $R_{a,b}^{(1)} < R_{\text{cut}} = 0.5$  nm are defined as being part of the same contact cluster. These aggregates include optical clusters (and through optical clusters, aligned clusters as well) and also aggregates in which the molecules interact through more promiscuous side chain–core and side chain–side chain interactions. Due to the fact that not all molecules in a contact cluster interact *via* their aromatic cores, contact clusters are not expected to display desirable optoelectronic properties.

### 3. Results and discussion

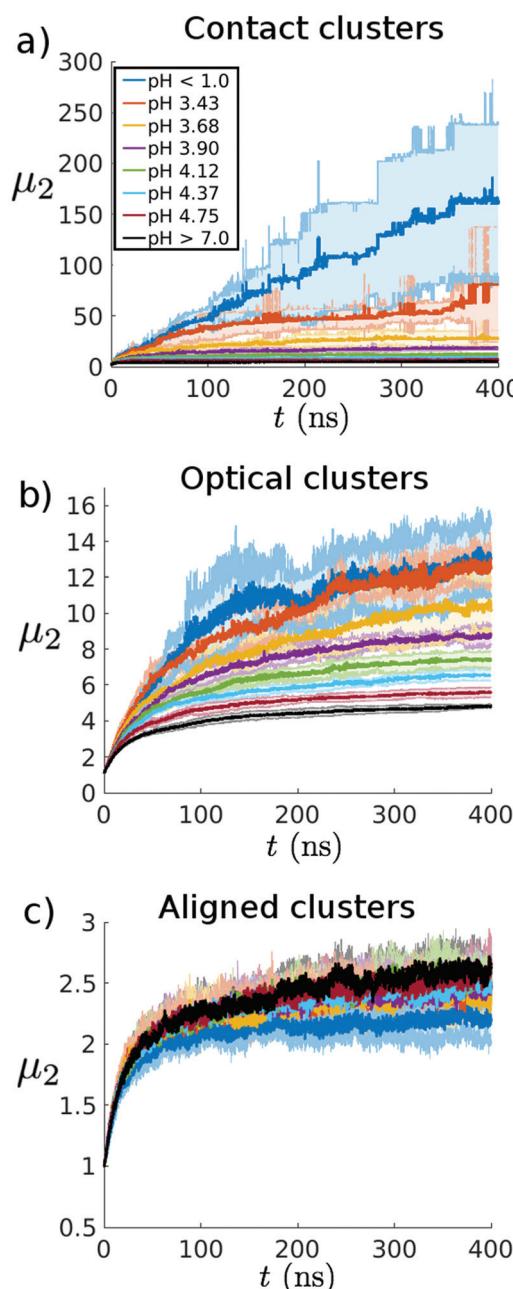
#### 3.1 Influence of pH on early-stage assembly

Experiment<sup>11,18–20</sup> and computation<sup>12,21</sup> have shown aggregation of DFAG-OPV3-GAFD monomers to proceed under low-pH conditions due to protonation of the terminal Asp residues and a concomitant elimination of intermolecular electrostatic repulsion. In this section, we investigate the effects of changing the percentage of charged molecules as a proxy for changing the pH to reveal a reduction in growth rate of optical and contact clusters with increasing pH. Conversely, the growth rate of aligned clusters is unaffected. Due to the hierarchical nature of aggregation (Fig. 2), elevated pH results in a slowdown in the aggregation rate, but the proportion of better-aligned aggregates in solution increases. We quantify the assembly kinetics using the Smoluchowski coagulation theory and develop a simple scaling model to understand the trends in the assembly kinetics with pH.

**3.1.1 Rate of growth of mass-averaged cluster size falls with increasing pH.** The instantaneous  $n$ -averaged cluster size – where we intend by cluster size the number of monomers in a cluster – is defined as,<sup>55</sup>

$$\mu_n(t) = \frac{\sum_{i=1}^N m_i^n N_i(t)}{\sum_{i=1}^N m_i^{(n-1)} N_i(t)} = \frac{\sum_{i=1}^N i^n N_i(t)}{\sum_{i=1}^N i^{(n-1)} N_i(t)}, \quad (9)$$

where  $m_i = m_{\text{mon}} i$  is the mass of a cluster containing  $i$  monomers of mass  $m_{\text{mon}}$  and  $N_i(t)$  is the total number of clusters of size  $i$  extant in the simulation at time  $t$ . In Fig. 3, we plot the mass-averaged cluster size  $\mu_2$  as a function of time for contact, optical, and aligned clusters at each of the eight pH values considered. The mass-averaged cluster size is an experimentally-measurable quantity that provides a good metric with which to track the aggregation state of the system and can be preferred over the number-averaged cluster size  $\mu_1$  due its



**Fig. 3** Temporal evolution of the mass-averaged cluster size  $\mu_2$  calculated at each of the eight pH values considered for (a) contact clusters, (b) optical clusters, and (c) aligned clusters. Different pH values are denoted by different colors.  $\mu_2(t)$  is computed as the mean over the five independent runs at each pH and represented as a solid line, and uncertainties represented by the shaded regions are estimated from the standard deviation taken over the five runs.

reduced weighting of the contributions of light clusters to the average.<sup>56</sup>

The trends in  $\mu_2(t)$  illustrate that increasing pH increasingly disfavors the formation of large clusters, thereby slowing the rate at which contact and optical clusters are produced (Fig. 3). At the length and time scales accessible to our calculations, it

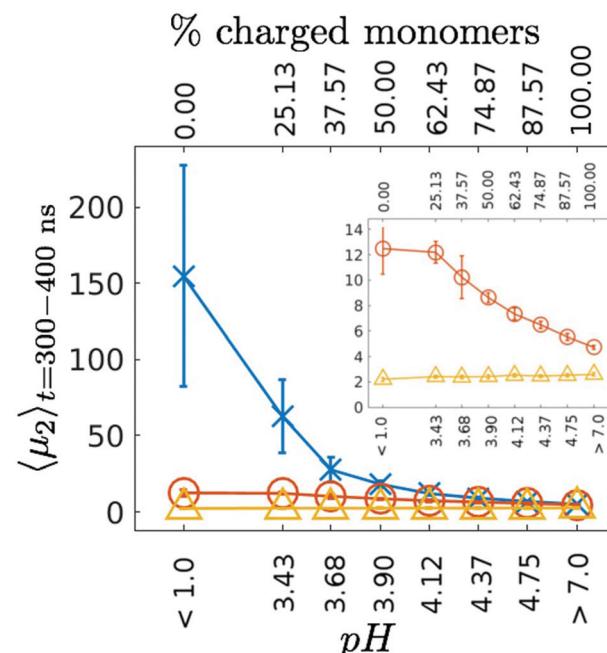
is difficult to determine from the high pH simulations ( $\text{pH} \geq 3.90$ ) whether the mass-averaged cluster sizes have attained stable values (*i.e.*, assembly has been arrested at a finite cluster size) or the growth rate is simply diminished to exceedingly low levels. A careful analysis of the slope of the growth curves as a function of time reveals that they indeed approach zero at long times (Fig. S1†), but within the uncertainty of our calculations and on such comparatively small scales it is not possible to conclusively state that growth has been completely arrested. Far longer and larger simulations would be required to determine whether there exists a critical pH above which contact and optical clusters are restricted to attain a maximum finite value, and below which they grow without bound to incorporate all monomers in the system. A body of computational and experimental work indicates that at sufficiently long time scales a fully protonated ( $\text{pH} < 1.0$ ) system will ultimately assemble into a single large contact cluster, and a fully deprotonated ( $\text{pH} > 7.0$ ) will form a distribution of small oligomers,<sup>11,12,18–21</sup> but the morphological state of systems at intermediate pH at very long times currently remains unknown. The predicted formation of small oligomeric aggregates even at high pH is consistent with dynamic light scattering measurements under these conditions.<sup>57</sup> Spectroscopy measurements, however, are consistent with free monomers, indicating that the limited assembly under high-pH conditions has relatively little impact on the spectroscopic properties of the system.<sup>19</sup> Theoretical resolution of the aggregate sizes and morphologies within which the spectroscopic response measurably differs from that of the free monomers would require quantum calculations employing a technique such as time-dependent density functional theory (TDDFT).<sup>58</sup>

In Fig. 4, we plot  $\langle \mu_2 \rangle_{t=300-400 \text{ ns}}$  as the mass-averaged cluster size at each pH averaged over the last 100 ns of all of the five independent runs at each pH. These data show that the mass-averaged cluster size decreases rapidly for contact clusters, moderately for optical clusters, and leaves aligned clusters unaffected. Furthermore, the mass-averaged cluster size under all three cluster definitions converges with increasing pH.

**3.1.2 Aggregate ordering increases with increasing pH.** The observed convergence of the cluster sizes defined by the three hierarchical metrics with increasing pH (Fig. 4) together with the hierarchical nesting of these metrics (*cf.* section 2.4) indicates that although the characteristic size of the contact and optical clusters falls with increasing pH, a larger fraction of the monomers within these clusters belong to aligned clusters with good ordering and in-register alignment between the aromatic cores. To quantify this trend, we define for each monomer  $i$  a ratio of the various cluster sizes within which it participates,

$$\zeta_i^{\alpha\beta} = \frac{r_i^\alpha}{r_i^\beta}, \quad (10)$$

where  $\alpha$  and  $\beta$  can take on one of three cluster signifiers, C for contact, O for optical, or A for aligned, and  $r_i^\alpha$  represents the size in monomers of the cluster of type  $\alpha$  to which monomer

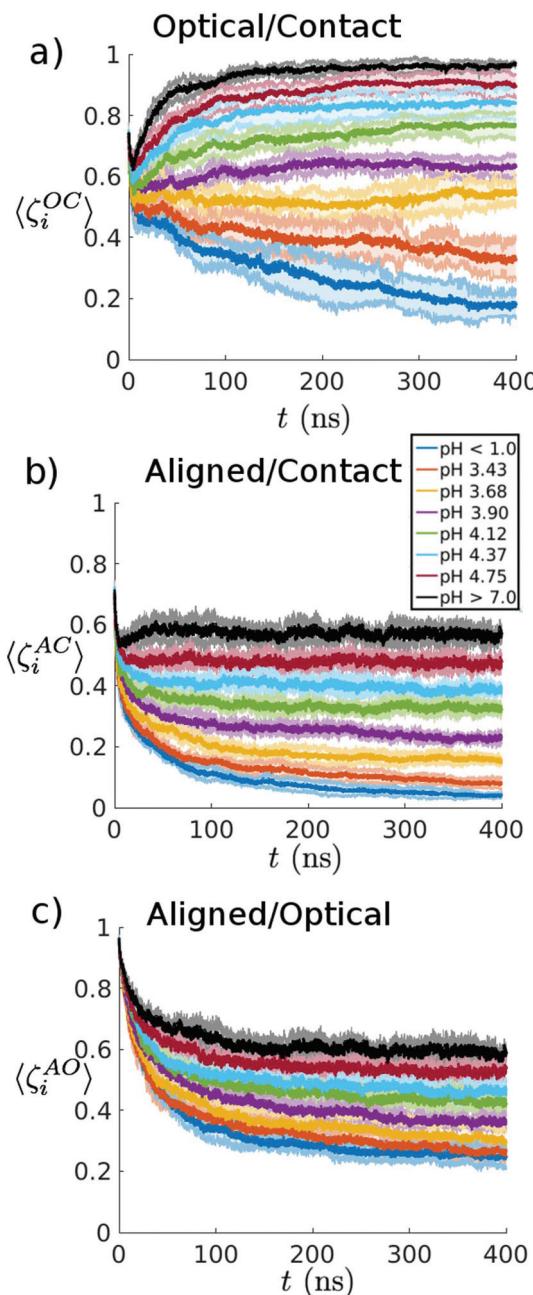


**Fig. 4** The mean of the mass-averaged cluster size taken over the final hundred ns of all five runs  $\langle \mu_2 \rangle_{t=300-400 \text{ ns}}$  for contact, optical, and aligned clusters at each of the eight pH values considered. Error bars are computed as the mean of the standard deviation over the final hundred ns. The inset shows a zoomed in view of the trends in the optical and aligned clusters. The pH has a strong effect on the mass-averaged cluster size of contact clusters, a moderate effect on that of optical clusters, and no effect on that of aligned clusters. Blue crosses represent contact clusters, red circles optical clusters, and yellow triangles aligned clusters.

$i$  belongs. Averaging over all monomers  $i$  in the system,  $\langle \zeta_i^{\alpha\beta} \rangle$  provides a measure of the degree of order defined by metric  $\alpha$  within clusters defined by metric  $\beta$ . For example,  $\langle \zeta_i^{\text{AC}} \rangle$  provides a measure of the extent to which contact clusters are populated by aligned clusters. If  $\langle \zeta_i^{\text{AC}} \rangle = 1$ , then all contact clusters in the system are entirely composed of aligned clusters, indicating a high degree of intra-aggregate ordering, whereas  $\langle \zeta_i^{\text{AC}} \rangle \approx 0$  reveals that the monomers constituting contact clusters do not assemble into aligned morphologies.

We plot in Fig. 5 the time evolution of  $\langle \zeta_i^{\text{OC}} \rangle$ ,  $\langle \zeta_i^{\text{AC}} \rangle$ , and  $\langle \zeta_i^{\text{AO}} \rangle$  for each of the eight pH values studied averaged over all monomers  $i$  and over the five independent runs. All three plots demonstrate a marked increase in the intra-cluster ordering with increasing pH. Specifically, in going from pH < 1.0 to pH > 7.0, the extent to which contact clusters are composed of optical clusters increases from ~20% to ~90% (Fig. 5a) and of aligned clusters from ~10% to ~60% (Fig. 5b). Similarly, the aligned cluster composition of optical clusters increases from ~30% to ~65% over the same pH range (Fig. 5c).

Overall, the effect of increasing pH on DFAG-OPV3-GAFD aggregation is to disfavor the formation of large aggregates by slowing the assembly kinetics. We quantify assembly rates as a function of pH in section 3.3. The aggregates that do form at higher pH tend to be more ordered, containing a larger frac-



**Fig. 5** Time evolution of (a)  $\langle \zeta_i^{OC} \rangle$ , (b)  $\langle \zeta_i^{AC} \rangle$ , and (c)  $\langle \zeta_i^{AO} \rangle$  at each of the eight pH values considered. Different pH values are denoted by different colors, averages taken over all monomers in each of the five independent simulations at each time point, and uncertainties represented by the shaded regions are estimated from the standard deviation taken over the five runs. Clusters become increasingly well ordered with increasing pH.

tion of monomers with contacting or in-register aligned  $\pi$ -conjugated cores that are expected to give rise to emergent optoelectronic properties; however, it should be noted that this effect is primarily dependent on pH through cluster size, as smaller clusters in general tend to be more highly ordered. In the next section, we study the impact of pH on cluster morphology and growth kinetics.

### 3.2 Morphology of charged clusters

In this section, we characterize aggregate morphology and structure as a function of pH. Our primary findings are (i) the overall aggregate morphology is largely insensitive to pH, (ii) the system is well-mixed in that the proportion of charged to total monomers within each aggregate containing more than  $\sim 10$  monomers is roughly equal to that in the overall solution, but there is a slight preferential partitioning of charged monomers into small cluster sizes, and (iii) charged monomers tend to be more extended than uncharged monomers and the charged termini partition to the surface of the self-assembled clusters.

**3.2.1 Cluster morphology is insensitive to pH.** In order to test the effect of pH on the geometric shape of the aggregates, we compute for each contact cluster at each time step the elements of the gyration tensor  $S$ ,<sup>12,59</sup>

$$S_{pq} = \frac{1}{2R^2} \sum_{i=1}^R \sum_{j=1}^R (r_p^{(i)} - r_p^{(j)}) (r_q^{(i)} - r_q^{(j)}), \quad (11)$$

where  $p$  and  $q$  are indices that run over the three Cartesian coordinates  $x, y, z$ ,  $i$  and  $j$  are indices that run over the beads of all peptides in the cluster, and there are  $R = 29r$  total beads in a cluster of size  $r$  since each monomer contains 29 coarse-grained beads (*cf.* Fig. 1b). The eigenvalues of  $S$  are the ordered squares of the principal moments of the aggregates,  $(\xi_1^2, \xi_2^2, \xi_3^2)$ , from which a number of geometric properties can be calculated, including the radius of gyration,

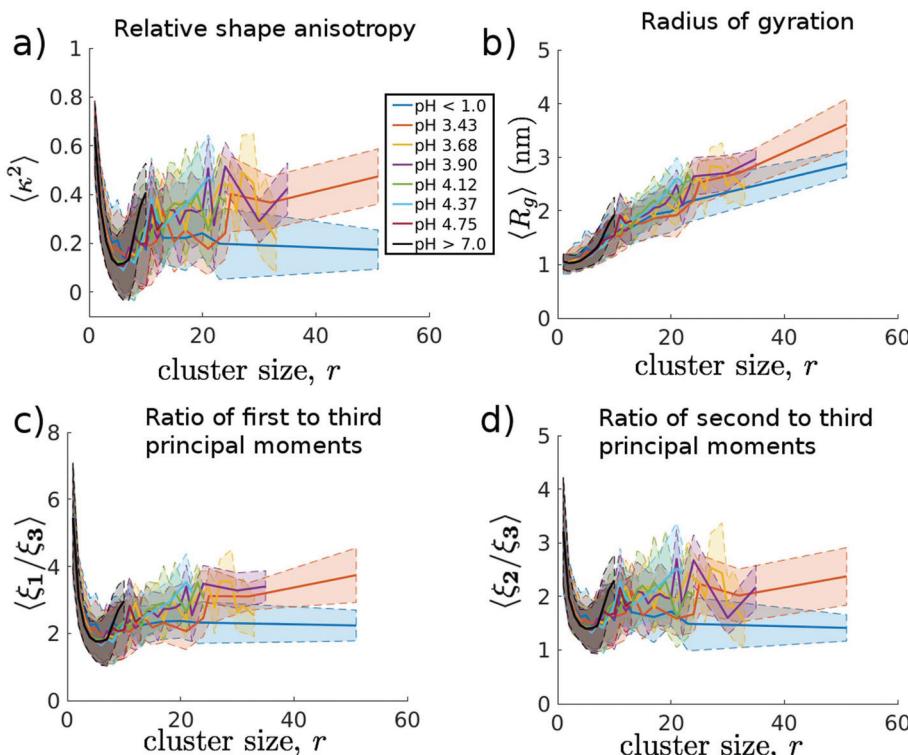
$$R_g = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}, \quad (12)$$

the relative shape anisotropy,

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

and the aspect ratios  $\xi_1/\xi_3$  and  $\xi_2/\xi_3$ . The radius of gyration is proportional to the root mean squared distance between the beads constituting the cluster, providing a measure of the compactness of the aggregate. The relative shape anisotropy ranges from zero to one and is a measure of aggregate symmetry. Aggregates with tetrahedral or higher-order symmetry, such as spherical aggregates, have a relative shape anisotropy of zero, whereas linear aggregates have a relative shape anisotropy of one.<sup>59</sup> The aspect ratios provide a measure of the relative elongation of the cluster along its principal axes.

In Fig. 6 we plot  $\kappa^2$ ,  $R_g$ ,  $\xi_1/\xi_3$ , and  $\xi_2/\xi_3$  as a function of contact cluster size  $r$  averaged over the entire time course of all five independent runs at each of the eight pH values considered. We consider clusters only of sizes  $r$  for which we possess sufficient statistics to report converged averages. All four morphological measures indicate that clusters adopt an elongated approximately ellipsoidal morphology, and their trends in cluster size as a function of pH are indistinguishable within error bars, with the exception of clusters of size  $r \gtrsim 20$  at  $pH < 1.0$  that tend to be more compact and spherical. Thus, for large enough cluster sizes, the presence of charged monomers



**Fig. 6** Cluster morphology as a function of contact cluster size  $r$  measured by the (a) relative shape anisotropy  $\kappa^2$ , (b) radius of gyration  $R_g$ , (c) ratio of the first and third principal moments  $\xi_1/\xi_3$ , and (d) ratio of the second and third principal moments  $\xi_2/\xi_3$ . Solid lines denote averages over all clusters of size  $r$  observed at any time in any of the five independent runs at each of the eight pH values considered. Different pH values are denoted by different colors. The shaded areas indicate standard deviations around the means. Only sizes  $r$  for which more than 1000 observations were recorded are plotted.

enhances the ellipsoidal nature of the aggregates, but the morphology is insensitive to the precise fraction of charged peptides over the 25–100% range studied in this work (see Table 1).

**3.2.2 pH has a mild effect on monomer orientations within aligned clusters.** To address the effect of pH upon the distribution of monomer orientations within aligned clusters, we computed for each monomer  $i$  in an aligned cluster the angle  $\Delta\theta_i$  formed by the unit vector  $\mathbf{v}_i$  pointing between the two distal aromatic rings in its  $\pi$ -conjugated core, and the mean such vector of the aligned cluster to which the monomer belongs  $\bar{\mathbf{v}} = \frac{1}{n} \sum_{k=1}^n \mathbf{v}_k$ . All  $\mathbf{v}_i$  vectors within an aligned cluster are

oriented to possess the same (arbitrary) sense in order to preclude trivial cancellations in the calculation of their mean. A 2-sample Kolmogorov-Smirnov test at a significance level of  $\alpha = 0.05$  permits us to reject the null hypothesis that the  $\Delta\theta$  distribution is invariant with pH, and we report the pH-dependent mean and standard deviations in Table 3. Monomers are generally quite tightly distributed around the mean orientation of the aligned cluster, with  $\langle \Delta\theta \rangle \approx 20\text{--}25^\circ$  and  $\sigma(\Delta\theta) \approx 12\text{--}13^\circ$ . Larger values of  $\langle \Delta\theta \rangle$  are observed at high pH where there are an increased fraction of charged monomers in the system possessing Coulombic repulsions between their charged termini. In sum, aligned clusters tend to form  $\beta$ -sheet like nanoaggregates with relatively tight orientational alignment of the constituent monomers that decreases slightly with increasing pH.

**Table 3** Dependence upon pH of the distribution of monomer orientations within aligned clusters around the cluster mean. We report the mean and standard deviation of  $\Delta\theta$  averaged over all monomers in aligned clusters observed in simulations conducted at each pH. There is a statistically significant pH dependence of the  $\Delta\theta$  distribution (2-sample Kolmogorov-Smirnov test;  $\alpha = 0.05$ ). The mean of  $\Delta\theta$  increases with pH in line with the increased fraction of charged monomers in the system and associated Coulombic repulsions between their charged termini

pH	$\langle \Delta\theta \rangle$ ( $^\circ$ )	$\sigma(\Delta\theta)$ ( $^\circ$ )
<1.0	20	12
3.43	21	13
3.68	21	13
3.90	22	12
4.12	23	13
4.37	23	13
4.75	24	13
>7.0	25	13

monomers in a contact cluster ( $N_q/N_{tot}$ )<sup>cluster</sup> as a function of cluster size  $r$ . We perform averaging over the entire time course of each of the five independent simulations conducted at each pH. We do not report data for the pH < 1.0 and pH > 7.0 systems that contain, respectively, exclusively protonated and exclusively deprotonated monomers.  $\langle(N_q/N_{tot})^{cluster}\rangle$  at  $r =$

1 is ~0.2 greater than the overall ratio of charged to uncharged monomers in the system ( $N_q/N_{tot}$ )<sup>system</sup> for all six pH values, indicating an elevated propensity for free monomers to be charged. Indeed  $\langle(N_q/N_{tot})^{cluster}\rangle$  exceeds  $(N_q/N_{tot})^{system}$  for  $r \lesssim 10$ , indicating that small aggregates overall tend to be more charged than average. Conversely, larger aggregates with  $r \gtrsim 10$  tend to possess a charged to total monomer ratio only slightly below the overall system composition.

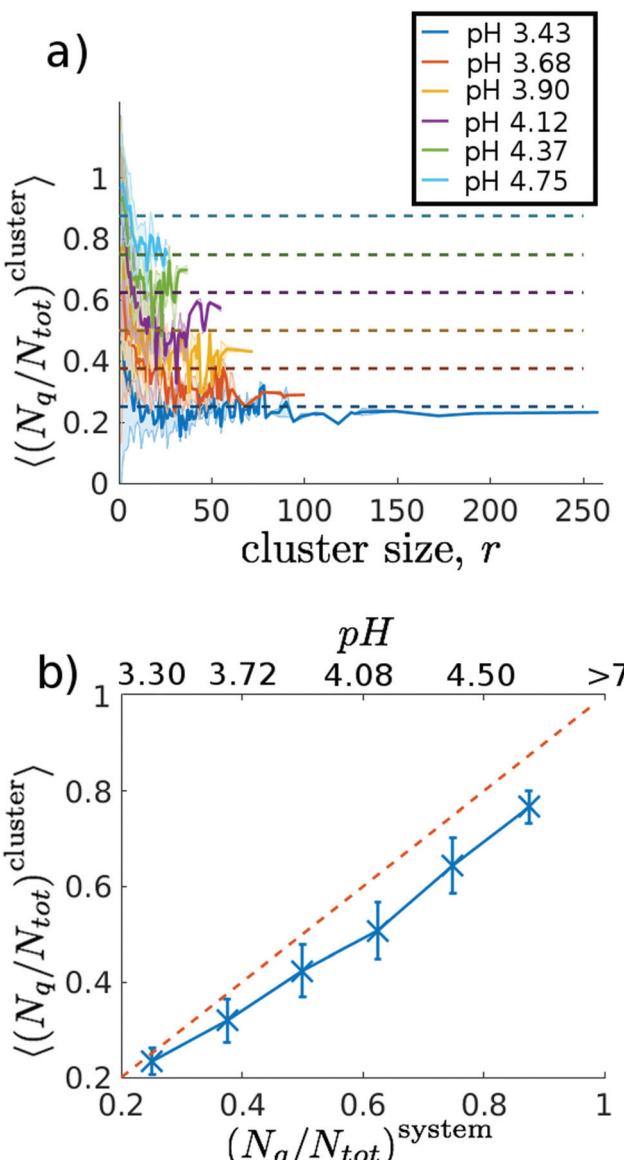
This observation is more clearly illustrated in Fig. 7b in which we average over all cluster sizes  $r > 10$  at each pH to construct a parity plot of  $\langle(N_q/N_{tot})^{cluster}\rangle$  against  $(N_q/N_{tot})^{system}$ . The data lie close to, but slightly below, the diagonal, indicating that larger clusters are of approximately the same composition as that of the system but slightly depleted in charged monomers due to the preferential partitioning of charged molecules into small contact clusters. Within any particular cluster there is no spatial dependence of the charged monomers such that, aside from the effect that we describe in the next section, clusters are approximately well-mixed.

**3.2.4 Charged termini lie further from the principal cluster axis than uncharged termini.** We now study the location of the termini of charged and uncharged monomers in contact clusters to determine if there are significant differences in their spatial arrangement. Considering the two coarse-grained beads containing the terminal Asp residue main chain C-termini, we compute the perpendicular distance of each terminal bead from the major axis of the cluster,

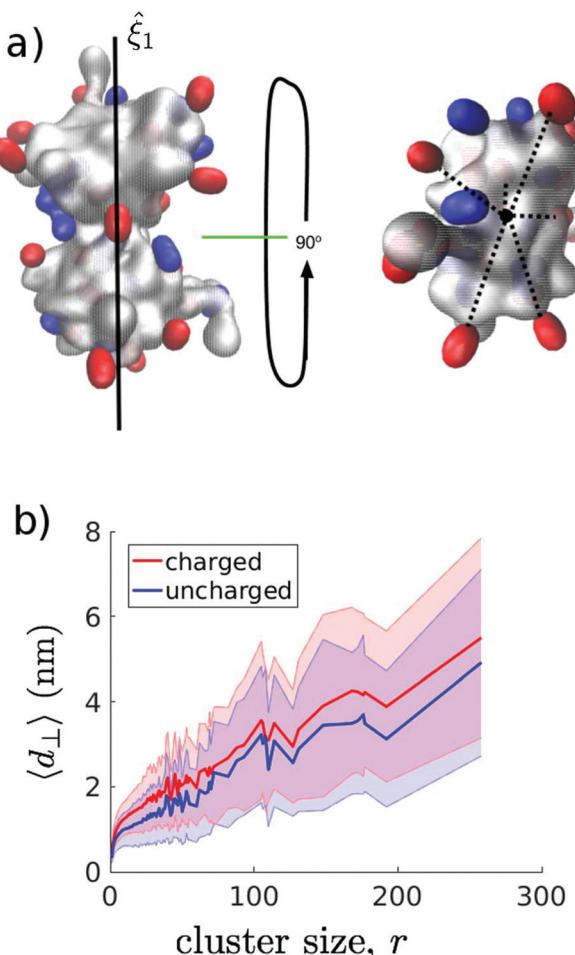
$$d_{\perp} = \left\| \mathbf{r}^b - \left( \mathbf{r}^b \cdot \hat{\xi}_1 \right) \hat{\xi}_1 \right\|, \quad (14)$$

where  $\mathbf{r}^b$  is the location of the bead in three-dimensional space in a coordinate frame with origin at the cluster center of mass, and  $\hat{\xi}_1$  is the unit vector coincident with the first principal eigenvector of the gyration tensor  $\mathbf{S}$  (see section 3.2.1 and eqn (11)). A schematic illustration of  $d_{\perp}$  for a representative cluster is presented in Fig. 8a.

The plot of  $d_{\perp}$  as a function of cluster size  $r$  in Fig. 8b illustrates that there is a small but measurable increase in the perpendicular distance of charged relative to uncharged termini. Although there is substantial variance in  $d_{\perp}$  for each cluster size, a one-sided Welch's  $t$ -test allows us to reject the null hypothesis that the distributions possess equal means ( $p < 0.05$ ) for over ninety percent of the  $r$  values at each pH. This supports the finding that charged monomers tend to be slightly more extended than uncharged monomers (see also section 3.2.5), and therefore the charged termini preferentially partition to the surface of contact clusters.  $d_{\perp}$  increases rapidly for small cluster sizes ( $r < 5$ ), then increases approximately linearly for larger aggregates. The increase of  $d_{\perp}$  beyond ~2.25 nm – half the linear extent of a fully-extended monomer – is primarily due to the contact clusters exhibiting branching behavior at large cluster sizes. We have previously reported this branching behavior and the formation of a fractal network under fully protonated (pH < 1.0) conditions,<sup>12</sup> and we provide representative snapshots of the cluster morphologies observed at each pH in Fig. 9. Complete movies of the simulation trajec-



**Fig. 7** Charged to total monomer ratio as a function of contact cluster size. (a) Ratio of charged to total monomers within contact clusters ( $N_q/N_{tot}$ )<sup>cluster</sup> as a function of contact cluster size  $r$ . Solid lines represent averages over the entire time course of each of the five independent simulations conducted at each pH, and the shaded area the standard deviation over these data. The dashed lines represent the overall ratio of charged to total monomers in the system ( $N_q/N_{tot}$ )<sup>system</sup>. (b) Parity plot of the ratio of charged to total monomers ( $N_q/N_{tot}$ )<sup>cluster</sup> averaged over all contact clusters of size  $r > 10$  to the overall system composition ( $N_q/N_{tot}$ )<sup>system</sup> at each pH. The red dotted line runs through the origin and has unit slope. Clusters of size  $r \lesssim 10$  tend to be enriched in charged monomers, whereas larger clusters have approximately the same charged to total ratio as the overall system.



**Fig. 8** Spatial arrangement of molecular termini within a contact cluster. (a) Schematic illustration of  $d_{\perp}$  for a representative contact cluster (see eqn (14)). The cluster is rendered as space-filling volume, where the grey surface indicates the spatial extent of the nonterminal portions of the aggregate upon which we have superposed red volumes indicating the location of charged termini and blue volumes the location of uncharged termini. This and all molecular snapshots and movies were rendered using the VMD molecular graphics software.<sup>25</sup> The two images presented are related by a 90° rotation into the plane of the page. In the image on the left, the solid black line represents the principal eigenvector of the gyration tensor  $\xi_1$ . In the image on the right, the solid black dot represents the principal eigenvector coming out of the page, and the dashed lines represent perpendicular distances  $d_{\perp}$  of selected charged termini. (b)  $d_{\perp}$  as a function of cluster size for charged and uncharged termini averaged over the course of all runs at all pH values. The shaded regions denote the standard deviation of the collected data. Charged termini tend to be located further from the central axis than uncharged termini (one-sided Welch's *t*-test,  $p < 0.05$ ), preferentially partitioning to the surface of the cluster.

stories from which these terminal frames were drawn are provided as Movies M1–8 in the ESI.†

**3.2.5 Charged monomers are more extended than uncharged monomers.** To complement the analysis of the previous section, we studied whether charged monomers in contact clusters tend to be more extended than uncharged monomers. In Fig. 10, we present the linear end-to-end dis-

tance  $d_{e2e}$  of the charged and uncharged monomers in contact clusters at each pH considered computed by averaging over the final 100 ns of the simulations. For comparison, we also present  $d_{e2e}$  for a single isolated charged monomer in solution and a single isolated uncharged monomer in solution computed from 30 ns equilibrated simulations of monomers solvated in a  $7 \times 7 \times 7$  nm<sup>3</sup> box.

At each pH considered, the charged monomers within contact clusters tend to be more extended than the uncharged monomers by  $\sim 0.5$  nm. A one-sided Welch's *t*-test allows us to reject the hypothesis that the means of the  $d_{e2e}$  distributions for the charged and uncharged monomers at each pH are equal ( $p < 0.01$ ). This provides additional support for the finding that charged monomers tend to be more extended than uncharged monomers due to electrostatic repulsion between the two termini, resulting in preferential segregation of charge to the exterior of the self-assembled aggregates. Furthermore, stacking of monomers within contact clusters induces an elongation of both charged and uncharged monomers relative to their mean extent in isolation. Finally, the mild decrease in the monomer extent within contact clusters with increasing pH is due to the decrease in large aggregates and concomitant increase in the numbers of small clusters and isolated monomers.

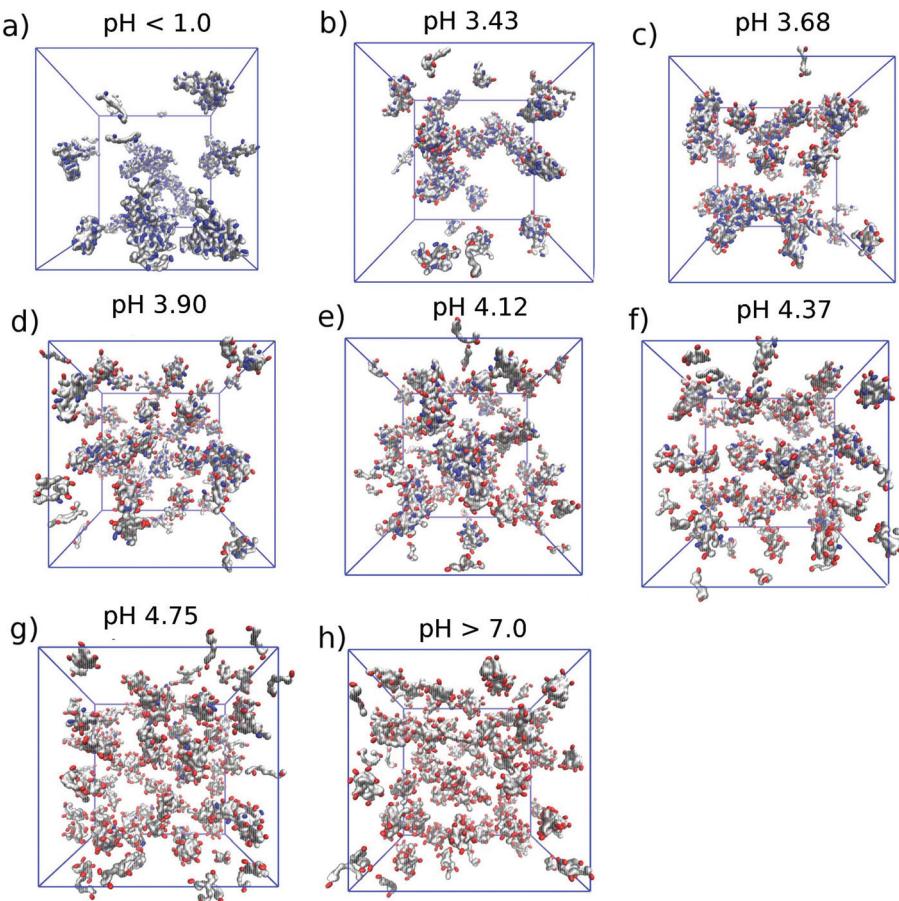
### 3.3 Modeling and explanation of growth kinetics

In this section we develop a quantitative model for pH-dependent aggregation, and appeal to the morphological results to understand the pH-dependence of the aggregation rate.

**3.3.1 Kinetics of growth can be modeled using the homogeneous Smoluchowski coagulation theory.** Following our previous study of self-assembly under fully protonated (pH < 1.0) conditions,<sup>12</sup> we use our simulation data to fit the pH-dependent parameters of a Smoluchowski coagulation model that enables us to describe the pH-dependence in the growth rates of different cluster types in a more quantitative way. The Smoluchowski coagulation model for the irreversible assembly of discrete aggregates is,<sup>60–62</sup>

$$\frac{dn_r}{dt} = \frac{1}{2} \sum_{i=1}^{r-1} K_{i,r-i} n_i(t) n_{r-i}(t) - \sum_{i=1}^{\infty} K_{r,i} n_r(t) n_i(t), \quad (15)$$

where  $n_r(t)$  is the number concentration of aggregates of size  $r$  at time  $t$  and  $K_{i,j}$  is a kernel for the second order rate constant of formation of aggregates of size  $i + j$  from coagulation between aggregates of sizes  $i$  and  $j$ . Analytical solutions of the Smoluchowski equations exist only for restricted classes of mathematical forms for the coagulation kernel.<sup>60,62–65</sup> In this work we adopt the homogeneity assumption for the kernels,  $K_{ai,aj} = a^2 K_{i,j}$ , that specifies a self-similar scaling for the second order rate constant with cluster size. In the analyses that follow, it is not necessary to specify the kernel further, with the homogeneous kernel providing a flexible model of the coagulation kinetics that admits analytical solutions for the moments of the cluster size distribution.<sup>62,63,66</sup> Although analytical expressions for the  $n_r(t)$  themselves are not available, knowledge of the moments is sufficient to predict the



**Fig. 9** Snapshots of the final  $t = 400$  ns frame of simulations conducted at pH (a)  $<1.0$ , (b) 3.43, (c) 3.68, (d) 3.90, (e) 4.12, (f) 4.37, (g) 4.75, and (h)  $>7.0$ . The molecules and clusters are rendered as space-filling volume, where the grey surface indicates the spatial extent of the nonterminal portions of the aggregate upon which we have superposed red volumes indicating the location of charged termini and blue volumes the location of uncharged termini. Coarse-grained water is omitted for clarity.

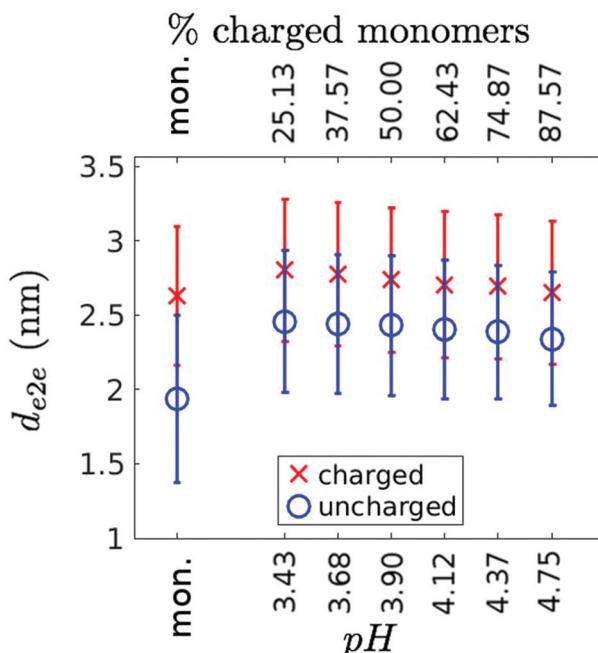
temporal evolution of the  $n$ -averaged cluster sizes  $\mu_n(t)$  (eqn (9)), and by fitting these analytical predictions to the simulation data we can regress kinetic parameters of the Smoluchowski model that optimally recapitulate the self-assembly dynamics observed in our coarse-grained molecular simulations. As discussed in section 3.1.1, it is unclear from the length and time scales accessible to our calculations whether the high-pH simulations exhibit kinetic arrest or just extremely slow aggregation kinetics. By developing a data-driven non-equilibrium kinetic model for the assembly dynamics, we permit the data to “speak for themselves” with the kinetically arrested state a special case of the model in which the coagulation rate constants tend to zero beyond some terminal cluster size.

Under the homogeneous kernel, the analytical solution for the time evolution of the mass-averaged cluster size for arbitrary initial conditions is,

$$\mu_2(t) = \mu_2(0) \left(1 + \frac{2t}{t_c}\right)^{\frac{1}{1-\lambda}}, \quad (16)$$

where  $\mu_2(0)$  is the initial mass-averaged cluster size and  $t_c$  the characteristic coagulation time.<sup>67</sup> Interpretable as the characteristic time for the productive collision of two monomers,<sup>63</sup> we make the approximation that  $t_c$  is independent of the monomer protonation state and therefore independent of the system pH. Consequently, we assume that all pH dependence of the assembly kinetics may be modeled through the scaling exponent  $\lambda$ . For  $\lambda > 0$  the aggregation rate constant increases with cluster size, for  $\lambda < 0$  it decreases, and for  $\lambda = 0$  we have a cluster size-independent rate constant that reduces to the special case of a constant kernel.

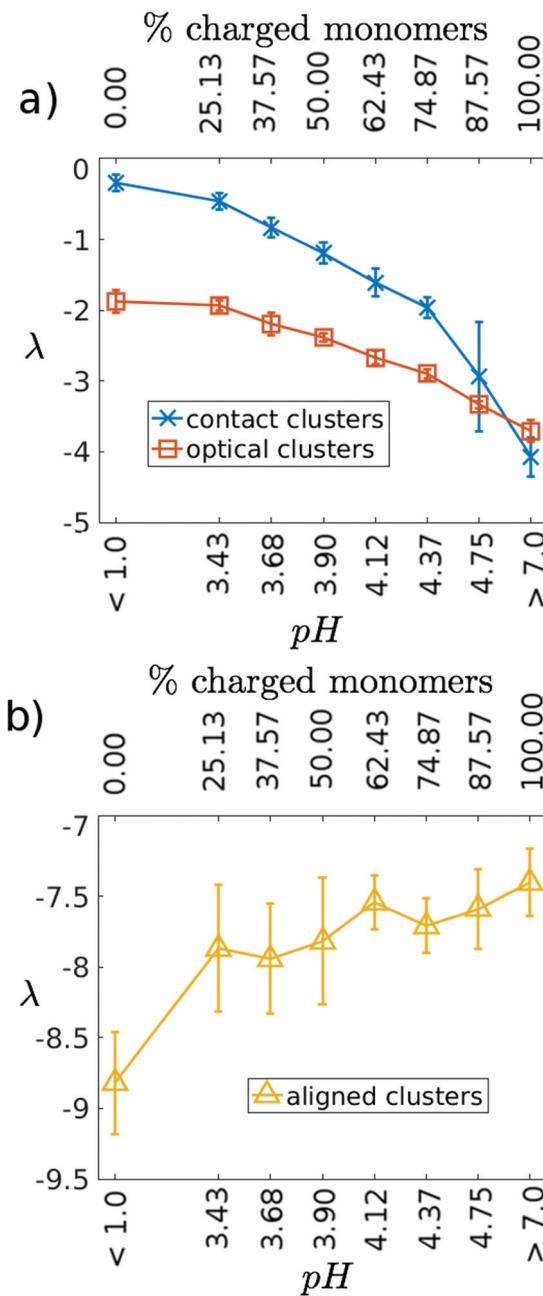
Considering contact, optical, and aligned clusters independently, we determine the optimal pH-independent value of  $t_c$  and pH-dependent  $\lambda$  for each cluster type by nonlinear least-squares fitting of the model predictions of the time evolution of the mass-averaged cluster size (eqn (16)) to our simulation data using the “fminbnd” and “fit” functions in MATLAB R2016a (The MathWorks, Inc., Natick, MA)<sup>68,69</sup> to a tolerance of  $1 \times 10^{-4}$ . The optimal estimates of the coagulation time emerging from this procedure are  $t_c^{\text{contact}} = (5.0198 \pm 0.0001)$  ns,  $t_c^{\text{optical}} = (0.5564 \pm 0.0001)$  ns, and  $t_c^{\text{aligned}} = (0.2377 \pm 0.0001)$



**Fig. 10** End to end distances  $d_{e2e}$  of charged and uncharged monomers in contact clusters. We present data for each of the six pH values considered (eliminating those that are exclusively composed of protonated or deprotonated monomers, as in section 3.2.3) computed by averaging over the final 100 ns of the five independent runs at each pH, and for isolated charged and uncharged monomers computed from 30 ns equilibrated simulations of single solvated molecules. Error bars correspond to standard deviations of the collected data around the mean. Both charged and uncharged monomers are more elongated within contact clusters than isolated in solution, and charged monomers are on average  $\sim 0.5$  nm more extended than uncharged monomers.

ns. The optimal estimates of the pH-dependent  $\lambda$  values are reported in Fig. 11. We present in Fig. S2–4 in the ESI† the Smoluchowski model predictions for  $\mu_2(t)$  employing the best-fit parameters for each of the three cluster types at each of the eight pH values considered. The good agreement of the model predictions with the simulation results provides *post hoc* validation that the Smoluchowski coagulation model with a homogeneous kernel provides a good description of the pH-dependent assembly kinetics.

Fig. 11a shows that  $\lambda$  for both contact and optical clusters decreases substantially with increasing pH. Appealing to the self-similar scaling of the homogeneous kernel  $K_{ai,aj} = a^2 K_{ij}$ , this trend indicates increasing suppression of the coagulation rate constants for the association of large clusters, and is consistent with the observed slowdown in the aggregation kinetics with increasing pH (see Fig. 3a and b). We also note that the pH  $< 1.0$  value of  $\lambda = -0.2 \pm 0.1$  for contact clusters is close to zero and therefore consistent with our previous work in which we demonstrated the constant kernel (*i.e.*,  $K_{ai,aj} = K_{ij} = K$ ) to provide a good description of the assembly kinetics in a fully protonated (uncharged) system. The convergence of the  $\lambda$  values for the contact and optical clusters at high pH indicates similar scaling behaviors in the coagulation rate constants as a



**Fig. 11** Best fit pH-dependent values of the homogeneous Smoluchowski kernel scaling exponent  $\lambda$  computed by nonlinear least-squares fitting of eqn (16) to the mass-averaged cluster size trajectories computed from our coarse-grained molecular simulations for (a) contact and optical, and (b) aligned clusters. Best fit values are computed as the mean over the five independent runs conducted at each pH, and error bars denote the calculated standard deviation. The decrease in  $\lambda$  with increasing pH for contact and optical clusters indicates increasing suppression of the coagulation rate constant for the aggregation of large clusters. With the exception of the datum for pH  $< 1.0$ ,  $\lambda$  for aligned clusters is relatively insensitive to pH.

function of cluster size. Together with the not too dissimilar values of  $t_c^{\text{contact}} = (5.0198 \pm 0.0001)$  ns and  $t_c^{\text{optical}} = (0.5564 \pm 0.0001)$  ns, this indicates that optical and contact cluster

assembly progresses approximately contemporaneously, underpinning both the high-pH convergence of the mass-averaged cluster size of contact and optical clusters (see Fig. 4) and the increasing proportion of each contact cluster that is composed of optical clusters (see Fig. 5).

Fig. 11b shows that with the exception of the datum at pH < 1.0,  $\lambda$  is approximately insensitive to pH lying within the range  $\lambda = [-8.5, -7.25]$ . A factor of two larger than the most negative  $\lambda$  value computed for the contact and optical clusters, this indicates that the suppression of large cluster aggregation rate constants is substantially greater for aligned clusters. This is consistent with the smaller size of aligned clusters (Fig. 3c) relative to contact and optical (Fig. 3a and b). That the pH < 1.0 lies outside the range spanned by the higher-pH data points may be a possible indication of finite size effects artificially suppressing the aligned cluster aggregation kinetics due to the rapid formation of optical and contact clusters under fully protonated conditions. A suite of simulations containing different numbers of monomers could be conducted to test this hypothesis.

**3.3.2 Phenomenological model of cluster growth.** We now develop a simple phenomenological model to help understand two features of the homogeneous kernel Smoluchowski model constructed in the previous section: why the contact cluster scaling exponent  $\lambda$  should be negative, and why it should become more negative with increasing pH. To do so we appeal to the three findings from the morphological characterization of contact clusters reported in section 3.2 that (i) contact cluster morphology is approximately ellipsoidal and largely insensitive to pH, (ii) clusters containing more than  $\sim 10$  monomers comprise a ratio of charged to total monomers approximately equal to the overall composition of the system, and (iii) the charged termini preferentially partition to the surface of the cluster. These results motivate a simple cluster model in which the fraction of charged monomers in the system  $f = 1/(1 + 10^{(pK_a - pH)})$  (see Table 1 and eqn (1)) may be used to estimate the number of charged monomers in a cluster containing  $N_{\text{tot}}$  total monomers as  $N_q \approx fN_{\text{tot}}$ . The volume of the cluster scales with  $N_{\text{tot}}$ , allowing us to rewrite this relation as  $N_q \sim f\ell^3$ , where  $\ell$  is a characteristic linear extent of the cluster. Since the termini of the charged monomers partition to the surface of the cluster, we may estimate the surface charge density as  $\sigma \sim N_q/\ell^2$ . Combining these two expressions, we find that  $\sigma \sim f\ell$  such that the charge density at the cluster surface increases linearly with both the fraction of charged monomers in the system and with the cluster linear extent. Accordingly, we should expect that the electrostatic repulsion between clusters should increase with both cluster size and pH due to the elevation of negative surface charge and reduction in the relative proportion of uncharged surface area available for productive aggregation. The increase in this electrostatic shielding with cluster size is the phenomenological root of the negative  $\lambda$  values for contact clusters reported in Fig. 11a, while its increase with the fraction of charged monomers in the system is the root of the downward trend in  $\lambda$  with increasing pH.

### 3.4 Influence of flow on early-stage self-assembly

We now investigate the role of flow on the alignment and morphology of clusters in the early stages of self-assembly.

**3.4.1 Flow has little effect on cluster morphology at short length and time scales.** To probe the effect of shear rate on the assembly kinetics we present in Fig. 12 the temporal evolution of the mass-averaged cluster size  $\mu_2$  for the contact, optical, and aligned clusters at the six shear rates of  $\dot{\gamma} = 0 \text{ s}^{-1}$ ,  $2.1 \times 10^4 \text{ s}^{-1}$ ,  $2.1 \times 10^5 \text{ s}^{-1}$ ,  $2.1 \times 10^6 \text{ s}^{-1}$ ,  $2.1 \times 10^7 \text{ s}^{-1}$ , and  $2.1 \times 10^8 \text{ s}^{-1}$ . We find the assembly rates of all three cluster types to be insensitive to shear rate on the time and length scales of our simulations. There is some evidence for an acceleration in the aggregation rate of contact clusters at the two largest shear rates considered ( $\dot{\gamma} = 2.1 \times 10^7 \text{ s}^{-1}$  and  $2.1 \times 10^8 \text{ s}^{-1}$ ) but the large error bars associated with sparse sampling of large cluster sizes at late times means that their trajectories cannot be distinguished within error bars from those at lower shear rates.

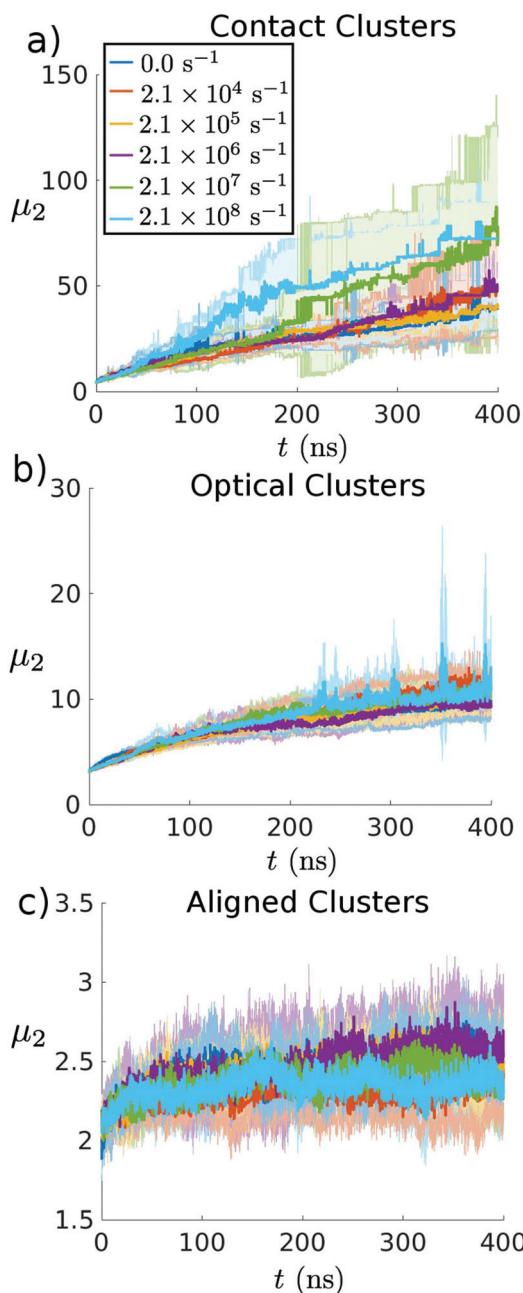
To explore the degree to which aggregates align with the flow, we compute for each contact cluster the angle it makes with the direction of the shear flow,

$$\cos(\theta_z) = \hat{\mathbf{n}}_z \cdot \hat{\xi}_1, \quad (17)$$

where  $\hat{\mathbf{n}}_z$  is a unit vector oriented perpendicular to the plane of constant flow velocity, and  $\hat{\xi}_1$  is the unit vector coincident with the first principal moment of the aggregate gyration tensor (see eqn (11)). Since the sense of  $\hat{\xi}_1$  is arbitrary, in performing this calculation we define its sense such that it forms an acute angle with  $\hat{\mathbf{n}}_z$ . If clusters perfectly align with the flow, then  $\hat{\mathbf{n}}_z$  and  $\hat{\xi}_1$  are orthogonal and an average over all clusters observed in our simulations will produce  $\langle \cos(\theta_z) \rangle = 0$ . If flow has no effect on cluster orientation and clusters adopt an isotropic distribution, then  $P(\theta_z) = \left( \int_0^{\pi/2} \sin(\theta_z) d\theta_z \right)^{-1}$  and  $\langle \cos(\theta_z) \rangle = \int_0^{\pi/2} \cos(\theta_z) P(\theta_z) \sin(\theta_z) d\theta_z = 0.5$ . Fig. 13 presents  $\langle \cos(\theta_z) \rangle$  as a function of cluster size at each of the six shear rates considered.

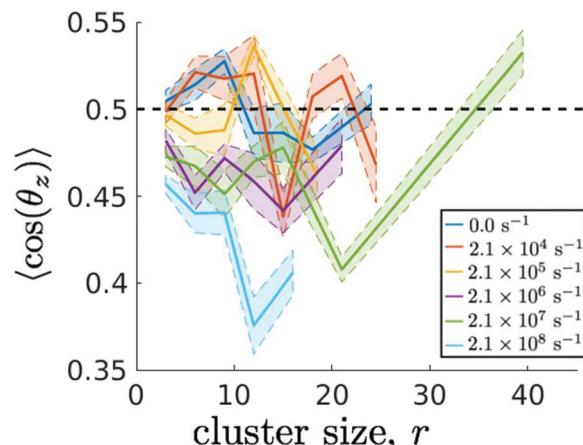
At none of the shear rates considered do we observe more than partial alignment over the time scale of our calculations. We observe the most significant deviation from an isotropic distribution of angles for the largest shear rate of  $\dot{\gamma} = 2.1 \times 10^8 \text{ s}^{-1}$ , at which  $\langle \cos(\theta_z) \rangle \approx 0.4$ , indicating some preferential alignment with flow. For the second and third largest shear rates of  $\dot{\gamma} = 2.1 \times 10^6 \text{ s}^{-1}$  and  $\dot{\gamma} = 2.1 \times 10^7 \text{ s}^{-1}$ ,  $\langle \cos(\theta_z) \rangle$  mainly lies slightly below 0.5, indicating a slight tendency towards alignment. At lower shear rates,  $\langle \cos(\theta_z) \rangle \approx 0.5$  indicating no preferential alignment with flow. Thus, large shear rates well beyond those realized in microfluidic devices ( $\dot{\gamma} \approx 2.5 \times 10^5 \text{ s}^{-1}$ ) are required to align the relatively small clusters formed during early-stage assembly.

To probe cluster morphology as a function of shear rate, in Fig. 14, we plot  $\kappa^2$ ,  $R_g$ ,  $\xi_1/\xi_3$ , and  $\xi_2/\xi_3$  (see section 3.2.1) as a function of contact cluster size  $r$  averaged over the entire time course of all five independent runs at each of the six shear rates. We retain only clusters of sizes  $r$  for which we possess sufficient statistics to report converged averages. All four mor-



**Fig. 12** Temporal evolution of the mass-averaged cluster size  $\mu_2$  calculated at each of the six shear rates considered for (a) contact clusters, (b) optical clusters, and (c) aligned clusters. Different shear rates are denoted by different colors.  $\mu_2(t)$  is computed as the mean over the five independent runs at each pH and represented as a solid line, and uncertainties represented by the shaded regions are estimated from the standard deviation taken over the five runs. Within error bars, flow has no effect on cluster growth rate on the length and time scales of our calculations, except to possibly accelerate contact cluster formation at the two highest shear rates considered.

phological measures indicate that clusters adopt an elongated approximately ellipsoidal morphology as they grow. Overlap of the data collected at each shear rate indicates that cluster morphology is generally insensitive to shear rate over the range of



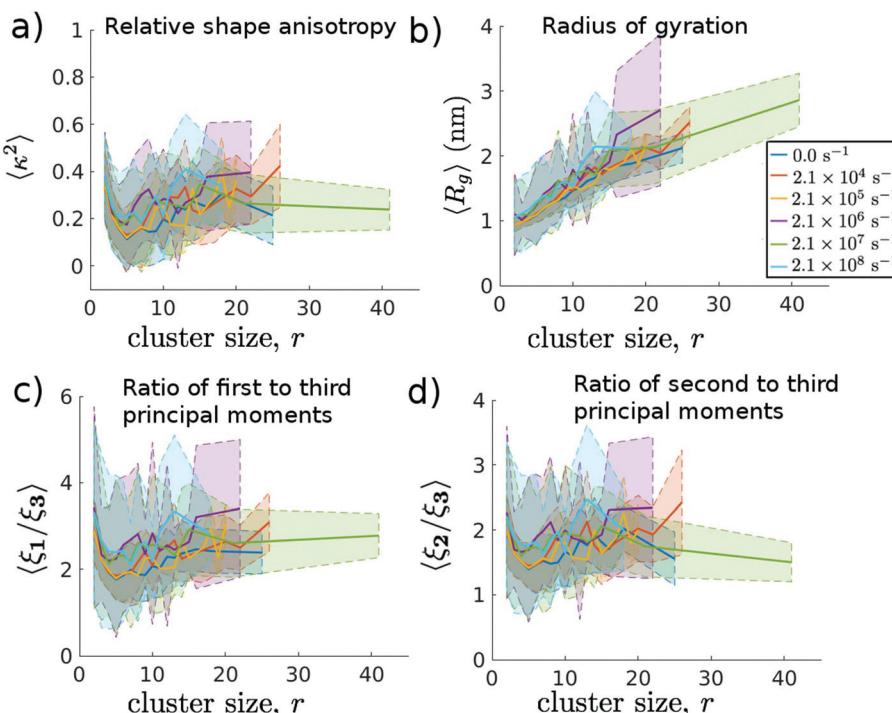
**Fig. 13** Alignment of contact clusters with the shear flow  $\langle \cos(\theta_z) \rangle$  as a function of cluster size  $r$ . Averages are taken over all clusters of size  $r$  observed at any time in any of the five independent runs at each of the six shear rates considered. Data is shown for all cluster sizes for which at least 1000 observations were recorded. Different shear rates are denoted by different colors. The solid lines are computed from a sliding window average that averages over three contiguous cluster sizes in order to smooth the data and evince its trends. The shaded areas indicate standard errors in the mean computed by randomly splitting the data for each cluster size into five equally sized blocks and computing the standard deviation over the blocks. The standard errors are subjected to the same window smoothing procedure as the means. Perfect alignment of the principal cluster axes with the shear flow produces  $\langle \cos(\theta_z) \rangle = 0$ , whereas an isotropic orientational distribution produces  $\langle \cos(\theta_z) \rangle = 0.5$  (indicated by the dashed black line).

shear flows and cluster sizes considered in this work. The mean and standard deviation of the distribution of monomer orientations within aligned clusters around the cluster mean (section 3.2.2) is  $\Delta\theta = (22 \pm 13)^\circ$  at all shear rates considered, and we have verified that there is no statistically significant dependence of the distribution on shear rate (2-sample Kolmogorov-Smirnov test;  $\alpha = 0.05$ ).

In sum, the effect of flow on cluster morphology in the early stages of assembly is relatively mild. At shear rates  $\dot{\gamma} \leq 2.1 \times 10^5 \text{ s}^{-1}$ , the cluster morphology, cluster alignment with the flow field, and rate of self-assembly are largely indistinguishable from quiescent conditions. Only at the very high shear rates of  $\dot{\gamma} \geq 2.1 \times 10^6 \text{ s}^{-1}$  do we find evidence of mild cluster alignment with the shear flow, and only for the even higher shear rates of  $\dot{\gamma} \geq 2.1 \times 10^7 \text{ s}^{-1}$  do we observe potentially accelerated rates of contact clusters.

### 3.5 Dimensionless analysis of flow in early-stage assembly

We can understand and quantify the relatively weak influence of flow on early-stage assembly by performing a dimensionless analysis inspired by Pipkin diagrams used in rheology<sup>22,23</sup> in which we construct ratios of the characteristic time scales for shear (the reciprocal shear rate  $\dot{\gamma}^{-1}$ ), rotational diffusion (the reciprocal rotational diffusivity  $D_r^{-1}$ ), and observation (the simulation time  $t_{\text{obs}}$ ). The shear rate  $\dot{\gamma}$  and simulation time  $t_{\text{obs}} = 400 \text{ ns}$  are determined from our simulation parameters, and



**Fig. 14** Cluster morphology as a function of contact cluster size  $r$  measured by the (a) relative shape anisotropy  $\kappa^2$ , (b) radius of gyration  $R_g$ , (c) ratio of the first and third principal moments  $\xi_1/\xi_3$ , and (d) ratio of the second and third principal moments  $\xi_2/\xi_3$ . The solid lines are averages over all clusters of size  $r$  observed at any time in any of the five independent runs at each of the six shear rates considered. Different shear rates are denoted by different colors. The shaded areas indicate standard deviations around the means. Only cluster sizes  $r$  for which more than 1000 observations were recorded are plotted.

we estimate  $D_r$  for 70 representative contact clusters comprising 6–330 monomers by analyzing their rotational dynamics as detailed in section 2.3. Given these three time scales, we form two independent dimensionless groups: the rotational Peclet number  $Pe_r$  and the Deborah number  $De$ .

The rotational Peclet number of a cluster,<sup>70</sup>

$$Pe_r = \frac{\dot{\gamma}}{D_r}, \quad (18)$$

is defined as the ratio of the shear rate  $\dot{\gamma}$  and the rotational diffusivity of the cluster  $D_r$ .  $Pe_r$  measures the relative magnitude of the hydrodynamic convection rate to the diffusion rate. Applied to the self-assembling contact clusters in our simulation, if  $Pe_r \gg 1$  then the strength of the shear flow dominates Brownian rotational diffusion and we should expect the clusters to align with the flow, whereas if  $Pe_r \ll 1$  then the shear rate is insufficient to dominate the rotational diffusion and we should expect to see an isotropic distribution of cluster orientations.<sup>11,71</sup>

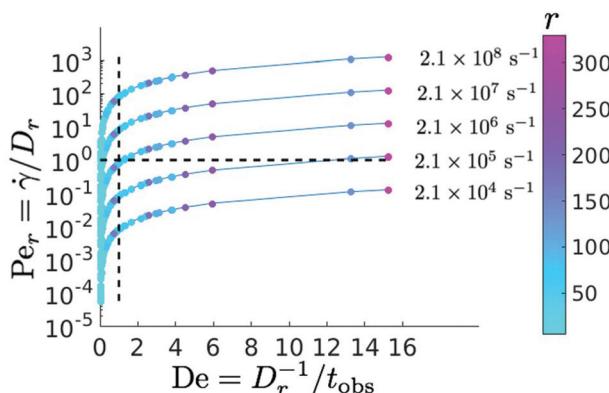
The Deborah number of a cluster,<sup>72–74</sup>

$$De = \frac{D_r^{-1}}{t_{obs}}, \quad (19)$$

is defined as the ratio of the characteristic time scale for rotational diffusion  $D_r^{-1}$  and the simulation time scale  $t_{obs}$ .  $De$

provides a measure of the time required for clusters to rotationally diffuse relative to the time over which we observe the system dynamics. For  $De \ll 1$  the time scale on which we observe the system far exceeds that over which rotational diffusion occurs and we should expect the clusters to comprehensively sample all of their accessible orientational states. For  $De \gg 1$  our observation time is far smaller than the rotational diffusion time scale and our simulations are too short to observe the orientational dynamics.

We present in Fig. 15 our Pipkin diagram showing the relationship between  $Pe_r$  and  $De$  at each of the five non-zero shear rates  $\dot{\gamma}$ . Each curve is parameterized by the rotational diffusivity  $D_r$  and we have superposed on these curves data points for each of the 70 representative contact clusters comprising 6–330 monomers for which  $D_r$  was computed. The dashed horizontal line at  $Pe_r = 1$  demarcates the turnover from the diffusive ( $Pe_r \ll 1$ ) to convective ( $Pe_r \gg 1$ ) regime. The dashed vertical line at  $De = 1$  delimits the transition from the regime of rapid ( $De \ll 1$ ) to slow ( $De \gg 1$ ) diffusion relative to the simulation time. Below the horizontal line the shear rate is insufficiently strong to align the clusters with the flow, and to the right of the vertical line the simulation time scale is too short to observe orientational relaxation. Accordingly, we should only expect to be able to observe the influence of flow within the upper left quadrant where the shear rate is sufficiently strong to impose alignment and the rotational



**Fig. 15** Pipkin diagram illustrating the relationship between rotational Peclét number  $\text{Pe}_r = \dot{\gamma}/D_r$ , measuring the relative rate of convective shear to rotational diffusivity, and Deborah number  $\text{De} = D_r^{-1}/t_{\text{obs}}$ , measuring the ratio of the diffusion time to the simulation time. The five curves correspond to each of the five non-zero shear rates upon which we have superposed the 70 points corresponding to the representative contact clusters for which rotational diffusivities were empirically computed. Each point is colored by cluster size. The dashed horizontal line at  $\text{Pe}_r = 1$  demarcates the turnover from the diffusive to convective regime, and the dashed vertical line at  $\text{De} = 1$  delimits the transition from the regime of rapid to slow diffusion relative to the simulation time. Only in the upper left quadrant is the shear flow sufficiently strong and diffusion rate sufficiently high that we should expect to observe cluster alignment with flow.

diffusivity sufficiently fast for us to observe alignment over the course of the simulation.

At the two lowest shear rates  $\dot{\gamma} = 2.1 \times 10^4 \text{ s}^{-1}$  and  $2.1 \times 10^5 \text{ s}^{-1}$ , almost all clusters containing up to  $r \approx 300$  monomers remain in the diffusive regime where the shear rate is insufficient to induce cluster alignment with flow and therefore couple to and influence the self-assembly morphology and dynamics. These shear rates are on the order of what has thus far been attained for this system in microfluidic cells,<sup>11</sup> and so we anticipate that flow should only become important in influencing the morphology and orientation of aggregates containing hundreds or thousands of monomers. At  $\dot{\gamma} = 2.1 \times 10^6 \text{ s}^{-1}$ , clusters of size  $r \gtrsim 100$  are sufficiently large to enter the convective regime, but do so to the right of the  $\text{De} = 1$  line. Accordingly, we should not expect to observe alignment of these aggregates over the course of our 400 ns simulations, and that alignment of these clusters with flow would occur in excess of microsecond time scales. For the highest shear rate  $2.1 \times 10^8 \text{ s}^{-1}$ , nearly all clusters of size  $r \lesssim 100$  fall into the upper left quadrant, while for the second highest shear rate  $\dot{\gamma} = 2.1 \times 10^7 \text{ s}^{-1}$ , most clusters between  $50 \lesssim r \lesssim 100$  and some smaller clusters do fall into the upper left quadrant, suggesting that we should observe alignment of these clusters with the flow. This is consistent with the observations that only for the highest shear rate did we observe significant alignment of clusters with the flow field (Fig. 13), and only for the two highest shear rates did we observe evidence for acceleration of the assembly rate (Fig. 12).

## 4 Conclusions and future work

Using a combination of coarse-grained molecular dynamics simulations and analytical theory, we have characterized the effects of pH and shear flow on the aggregation dynamics and morphologies of self-assembling  $\pi$ -conjugated DFAG-OPV3-GAFD peptides on time scales of hundreds of nanoseconds and length scales of hundreds of monomers. Overall, we have gained a molecular-level understanding of the effects of pH and flow on assembly and laid the groundwork to rationally manipulate pH and flow to engineer self-assembled aggregates with desirable optoelectronic properties.

We have shown that increasing pH strongly affects the assembly kinetics by disfavoring the formation of large aggregates due to a partitioning of charged residues to the aggregate surface and an increase in surface charge density that scales approximately linearly with cluster size and fraction of charged monomers in the system. The slow down in the kinetics with increasing pH is well modeled by Smoluchowski coagulation theory employing a homogeneous kernel where we have inferred kinetic parameters by fitting the model predictions to our simulation data. These analytical kinetic models provide a means to extrapolate predictions of the assembly kinetics to length and time scales far beyond those accessible to molecular simulation. The slow down in the assembly rate with elevated pH is accompanied by an increase in the alignment and ordering of the monomers within the aggregates. Our results suggest that pH can serve as a useful control variable to engineer the assembly of ordered aggregates. Specifically, we propose that conducting assembly under a gently decreasing pH ramp may promote long-range ordering within large aggregates *via* a hierarchical assembly mechanism wherein well-ordered smaller aggregates progressively coalesce into larger fibers as the surface charge density is reduced under increasingly acidic conditions to gradually promote further assembly. Experimental support for this conjecture is provided in recent work by Li *et al.*, who showed that reduced acidification rates achieved by introducing an organic layer to control acid diffusion produced slower assembly kinetics and more homogeneous self-assembled aggregates at length scales of hundreds of nanometers and time scales of hours.<sup>13</sup> These results are consistent with earlier work by Adams *et al.*, who showed that self-assembling fluorenylmethoxycarbonyl-dipeptides produced more homogeneous self-supporting hydrogels when acidified by the slow and controllable hydrolysis of glucono- $\delta$ -lactone into gluconic acid.<sup>75</sup> Nevertheless, no change in the final absorbance and emission spectra of the resultant aggregates was observed,<sup>13</sup> highlighting a need for future modeling work to probe the morphological effects of pH at longer length scales using ultra coarse-grained models, and the connection between the morphologies of aligned, optical, and contact clusters and their corresponding optical properties through electronic structure calculations. We also propose to conduct coarse-grained molecular dynamics simulations with time-varying pH to directly probe the microscopic mechanisms underpinning this hypothesized hierarchical aggregation

process. In doing so, we propose to devise programs of pH control – including potentially non-monotonic oscillatory pH schedules to aid in the healing of defects – to control the large-scale morphologies of these self-assembling systems and maximize their order and corresponding optoelectronic properties.

Our analysis of the impact of shear flow upon assembly morphology and kinetics reveals that flow only begins to couple to early-stage aggregation at very high shear rates in excess of  $\dot{\gamma} = 2.1 \times 10^6 \text{ s}^{-1}$ . Under these conditions, the shear flow induces alignment of clusters within the flow and accelerates the assembly kinetics. At experimentally realized shear rates within microfluidic devices, our Pipkin-inspired dimensionless analysis suggests that the flow field is expected to influence the orientation and aggregation of clusters comprising in excess of hundreds of monomers over hundreds of nanosecond time scales or longer. The fact that cluster size increases approximately monotonically even at very high shear rates shows that flow does not induce fragmentation of clusters containing up to  $\sim 180$  monomers, but we propose that it may disturb the branched nature of very large contact clusters that we previously demonstrated to adopt a fractal network architecture of dimensionality  $\sim 1.5$ .<sup>12</sup> These findings suggest that the experimentally verified role of flow in enhancing optoelectronic properties of these self-assembled aggregates<sup>11,76,77</sup> is due to the flow-induced alignment and ordering of aggregates to form elongated fibrils and fibers at time and length scales exceeding hundreds of ns and tens of nm. To explicitly probe this hypothesis, we propose that it would be instructive in future work to perform simulations of the effect of flow on the alignment and morphology of preassembled aggregates comprising hundreds of monomers harvested from the calculations performed in this work.

In future work, we plan to perform quantum calculations employing time-dependent density functional theory (TDDFT)<sup>58</sup> to characterize the optoelectronic properties and underlying electronic mechanisms within archetypal aggregates observed in our classical molecular dynamics simulations and thereby directly quantify the role of morphology and ordering upon the emergent optical and electronic responses. We also propose to develop a bottom-up ultra coarse-grained model for peptide assembly<sup>78</sup> to reach exceedingly long time and length scales and ultimately attempt to make contact with experimentally-accessible time and length scales. Finally, we also plan to consider different chemistries in the DXXX-II-XXXD series to probe the effects of chemistry on aggregation and optoelectronic properties.

## Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award # DE-SC0011847. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of

Science of the U.S. Department of Energy under Contract no. DE-AC02-05CH11231. R. A. M. acknowledges a graduate fellowship from the Computational Science and Engineering Program at the University of Illinois. We thank Prof. Charles Schroeder, Dr Bo Li, and Mr Lawrence Valverde for fruitful discussions.

## References

- X. Guo, M. Baumgarten and K. Müllen, *Prog. Polym. Sci.*, 2013, **38**, 1832–1908.
- L. Bian, E. Zhu, J. Tang, W. Tang and F. Zhang, *Prog. Polym. Sci.*, 2012, **37**, 1292–1331.
- S. H. Kim and J. R. Parquette, *Nanoscale*, 2012, **4**, 6940–6947.
- F. J. M. Hoeben, P. Jonkheijm, E. W. Meijer and A. P. H. J. Schenning, *Chem. Rev.*, 2005, **105**, 1491–1546.
- D. Pinotsi, L. Grisanti, P. Mahou, R. Gebauer, C. F. Kaminski, A. Hassanali and G. S. Kaminski Schierle, *J. Am. Chem. Soc.*, 2016, **138**, 3046–3057.
- A. Shukla, S. Mukherjee, S. Sharma, V. Agrawal, K. R. Kishan and P. Guptasarma, *Arch. Biochem. Biophys.*, 2004, **428**, 144–153.
- A. Handelman, N. Kuritz, A. Natan and G. Rosenman, *Langmuir*, 2015, **32**, 2847–2862.
- D. W. P. M. Löwik, E. H. P. Leunissen, M. van den Heuvel, M. B. Hansen and J. C. M. van Hest, *Chem. Soc. Rev.*, 2010, **39**, 3394–3412.
- R. V. Ulijn and A. M. Smith, *Chem. Soc. Rev.*, 2008, **37**, 664–675.
- M. Zelzer and R. V. Ulijn, *Chem. Soc. Rev.*, 2010, **39**, 3351–3357.
- A. B. Marciel, M. Tanyeri, B. D. Wall, J. D. Tovar, C. M. Schroeder and W. L. Wilson, *Adv. Mater.*, 2013, **25**, 6398–6404.
- R. A. Mansbach and A. L. Ferguson, *J. Phys. Chem. B*, 2017, **121**, 1684–1706.
- B. Li, S. Li, Y. Zhou, H. A. M. Ardoña, L. R. Valverde, W. L. Wilson, J. D. Tovar and C. M. Schroeder, *ACS Appl. Mater. Interfaces*, 2017, **9**, 3977–3984.
- L. Brunsved, B. J. B. Folmer, E. W. Meijer and R. P. Sijbesma, *Chem. Rev.*, 2001, **101**, 4071–4098.
- S. Fleming and R. V. Ulijn, *Chem. Soc. Rev.*, 2014, **43**, 8150–8177.
- M. J. Webber, E. A. Appel, E. Meijer and R. Langer, *Nat. Mater.*, 2016, **15**, 13–26.
- H. Cui, M. J. Webber and S. I. Stupp, *Pept. Sci.*, 2010, **94**, 1–18.
- B. D. Wall and J. D. Tovar, *Pure Appl. Chem.*, 2012, **84**, 1039–1045.
- B. D. Wall, Y. Zhou, S. Mei, H. A. M. Ardoña, A. L. Ferguson and J. D. Tovar, *Langmuir*, 2014, **30**, 11375–11385.
- B. D. Wall, A. E. Zacca, A. M. Sanders, W. L. Wilson, A. L. Ferguson and J. D. Tovar, *Langmuir*, 2014, **30**, 5946–5956.

- 21 B. A. Thurston, J. D. Tovar and A. L. Ferguson, *Mol. Simul.*, 2016, **1**, 1–21.
- 22 A. Pipkin, *Lecture Notes in Viscoelasticity*, 1972.
- 23 R. H. Ewoldt, A. E. Hosoi and G. H. McKinley, *Annu. Trans. – Nord. Rheol. Soc.*, 2007, **15**, 3.
- 24 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
- 25 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 26 J. Mongan and D. A. Case, *Curr. Opin. Struct. Biol.*, 2005, **15**, 157–163.
- 27 M. S. Lee, F. R. Salsbury and C. L. Brooks, *Proteins: Struct., Funct., Bioinf.*, 2004, **56**, 738–752.
- 28 M. Dlugosz, J. M. Antosiewicz and A. D. Robertson, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 2004, **69**, 021915.
- 29 M. Dlugosz and J. M. Antosiewicz, *Chem. Phys.*, 2004, **302**, 161–170.
- 30 J. Mongan, D. A. Case and J. A. McCammon, *J. Comput. Chem.*, 2004, **25**, 2038–2048.
- 31 W. D. Bennett, A. W. Chen, S. Donnini, G. Groenhof and D. P. Tieleman, *Can. J. Chem.*, 2013, **91**, 839–846.
- 32 C. E. Housecroft and E. C. Constable, *Chemistry: An Introduction to Organic, Inorganic and Physical Chemistry*, Pearson education, Essex, UK, 4th edn, 2010.
- 33 D. Lu, Y. Zhang, Y. Li, C. Luo, X. Wang, X. Guan, H. Ma, X. Zhao, Q. Wei and Z. Lei, *J. Polym. Sci., Part A: Polym. Chem.*, 2017, **55**, 207–212.
- 34 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
- 35 L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.
- 36 D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman and S. J. Marrink, *J. Chem. Theory Comput.*, 2013, **9**, 687–697.
- 37 V. Rühle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, **5**, 3211–3223.
- 38 W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.
- 39 S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta and S. J. Marrink, *PLoS Comput. Biol.*, 2010, **6**, e1000810.
- 40 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 41 M. Parrinello, *J. Appl. Phys.*, 1981, **52**, 7182.
- 42 R. Hockney and J. Eastwood, *Computer Simulation Using Particles*, Taylor & Francis, 1988, p. 540.
- 43 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511.
- 44 T. Sodemann, B. Dünweg and K. Kremer, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 2003, **68**, 046702.
- 45 A. W. Lees and S. F. Edwards, *J. Phys. C: Solid State Phys.*, 1972, **5**, 1921–1928.
- 46 P. Kundu and I. Cohen, *Fluid Mechanics*, Elsevier Science, 2010.
- 47 M. Greenspan and C. E. Tschiegg, *J. Acoust. Soc. Am.*, 1959, **31**, 75–76.
- 48 S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.*, 2013, **42**, 6801–6822.
- 49 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 50 B.-Y. Cao and R.-Y. Dong, *J. Chem. Phys.*, 2014, **140**, 034703.
- 51 R.-Y. Dong and B.-Y. Cao, *Sci. Rep.*, 2014, **4**, 6120.
- 52 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic press, 2001, vol. 1.
- 53 *Asphaltenes and Asphalts*, 2, ed. T. F. Yen and G. Chilingarian, Elsevier, Amsterdam, 2000.
- 54 J. Wang and A. L. Ferguson, *J. Phys. Chem. B*, 2016, **120**, 8016–8035.
- 55 M. Rubinstein and R. Colby, *Polymer Physics*, OUP Oxford, 2003.
- 56 M. Sedghi, L. Goual, W. Welch and J. Kubelka, *J. Phys. Chem. B*, 2013, **117**, 5765–5776.
- 57 H. A. M. Ardoña and J. D. Tovar, *Chem. Sci.*, 2015, **6**, 1474–1484.
- 58 M. A. Marques and E. K. Gross, *Annu. Rev. Phys. Chem.*, 2004, **55**, 427–455.
- 59 D. N. Theodorou and U. W. Suter, *Macromolecules*, 1985, **18**, 1206–1214.
- 60 J. A. Wattis, *Phys. D*, 2006, **222**, 1–20.
- 61 M. Smoluchowski, *Z. Phys.*, 1916, **17**, 557–585.
- 62 F. Leyvraz, *Phys. Rep.*, 2003, **383**, 95–212.
- 63 A. Modler, K. Gast, G. Lutsch and G. Damaschun, *J. Mol. Biol.*, 2003, **325**, 135–148.
- 64 R. P. Treat, *J. Phys. A: Math. Gen.*, 1990, **23**, 3003–3016.
- 65 F. Leyvraz and H. R. Tschudi, *J. Phys. A: Math. Gen.*, 1981, **14**, 3389–3405.
- 66 B. J. Olivier and C. M. Sorensen, *Phys. Rev. A*, 1990, **41**, 2093–2100.
- 67 B. J. Olivier, C. M. Sorensen and T. W. Taylor, *Phys. Rev. A*, 1992, **45**, 5614–5623.
- 68 G. Forsythe, M. Malcolm and C. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, 1977.
- 69 R. Brent, *Algorithms for Minimization Without Derivatives*, Dover Publications, 2013.
- 70 C. Tannous, *Surf. Sci.*, 2011, **605**, 923–929.
- 71 R. G. Larson, *The Structure and Rheology of Complex Fluids*, Oxford University Press, New York, 1999.
- 72 M. Reiner, *Phys. Today*, 1964, **17**, 62.
- 73 J. Dealy, *Rheol. Bull.*, 2010, **79**, 14–18.
- 74 R. Poole, The British Society of Rheology, *Rheol. Bull.*, 2012, **53**, 32–39.
- 75 D. J. Adams, M. F. Butler, W. J. Frith, M. Kirkland, L. Mullen and P. Sanderson, *Soft Matter*, 2009, **5**, 1856–1862.
- 76 S. Zhang, M. A. Greenfield, A. Mata, L. C. Palmer, R. Bitton, J. R. Mantei, C. Aparicio, M. O. De La Cruz and S. I. Stupp, *Nat. Mater.*, 2010, **9**, 594–601.
- 77 B. D. Wall, S. R. Diegelmann, S. Zhang, T. J. Dawidczyk, W. L. Wilson, H. E. Katz, H.-Q. Mao and J. D. Tovar, *Adv. Mater.*, 2011, **23**, 5009–5014.
- 78 J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner and G. A. Voth, *J. Chem. Theory Comput.*, 2013, **9**, 2466–2480.