

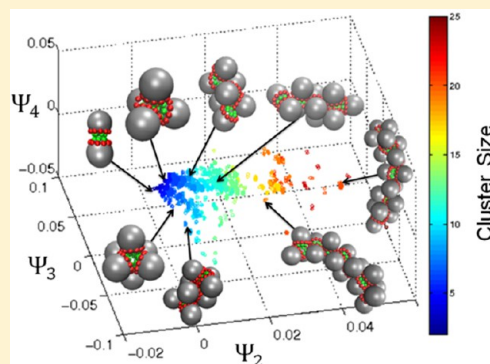
Nonlinear Machine Learning of Patchy Colloid Self-Assembly Pathways and Mechanisms

Andrew W. Long and Andrew L. Ferguson*

Department of Materials Science and Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

S Supporting Information

ABSTRACT: Bottom-up self-assembly offers a means to synthesize materials with desirable structural and functional properties that cannot easily be fabricated by other techniques. An improved understanding of the structural pathways and mechanisms by which self-assembling materials spontaneously form from their constituent building blocks is of value in understanding the fundamental principles of assembly and in guiding inverse building block design. We present an approach to infer systematically assembly pathways and mechanisms by nonlinear data mining of molecular simulation trajectories using diffusion maps. We have validated our methodology in applications to Brownian dynamics simulations of the assembly of anisotropic “patchy colloids” into polyhedral aggregates. For particles designed to form tetrahedral aggregates, we identify two divergent assembly pathways leading to chains of interlocking dimers and tetramers and chains of interlocking trigonal planar trimers. For particles designed to assemble icosahedral aggregates, our approach recovers two distinct assembly pathways corresponding to monomeric addition and budding from a disordered liquid phase. These assembly routes were previously reported by inspection of simulation trajectories by Wilber et al. (*J. Chem. Phys.* **2007**, 127, 085106; *J. Chem. Phys.* **2009**, 131, 175102), validating the capacity of our approach to systematically recover assembly mechanisms previously discernible only by trajectory visualization.



1. INTRODUCTION

Self-assembly is the process by which structured aggregates form by the spontaneous organization of their constituent building blocks.¹ In materials engineering, bottom-up self-assembly has been used to produce a range of materials with desirable structural and functional properties that cannot easily be fabricated by any other means.^{2–6} Materials of particular interest include microelectronics and optoelectronics,^{4,7–10} photonic crystals,¹¹ and nanobiomaterials for biomedical applications¹² including antimicrobial nanostructures,¹³ drug delivery vessels,^{14,15} and 3D cell scaffolds for regenerative tissue engineering.³ In natural systems, self-assembly underpins the assembly of not only structures vital to life processes such as protein complexes, actin filaments, and cell membranes^{4,5} but also pathological assemblies such as viral capsids^{16,17} and fibrillar aggregates implicated in Alzheimer’s and Huntington’s disease.^{18–20} An improved understanding of material assembly mechanisms is of fundamental scientific interest and great value in guiding the design of novel materials and development of new therapeutics. Nevertheless, it remains a significant challenge to infer the precise microscopic mechanisms and structural pathways by which building blocks assemble into their terminal aggregates.^{4,9,21,22}

Advances in particle patterning techniques have enabled the development of colloidal particles with tunable anisotropic surface interactions that can be programmed to self-assemble a variety of complex aggregates.^{4,23,24} These “patchy colloids” are

of interest in the development of 3D photonic crystals,^{23,25,26} biomimetic models of viral capsid assembly,^{27–29} and drug encapsulants.^{23,30–34} In conjunction with the development of new synthesis techniques, theory and computation have been used to understand aggregate formation and predict phase behavior^{35,36} and to develop design rules to optimize the stability of desired aggregates.^{9,37} Determining the thermodynamic stability of competing aggregates is a critical prerequisite to the understanding and design of self-assembly systems, but inference of the assembly pathways and mechanisms by which these aggregates assemble is necessary to fully characterize the underlying kinetic assembly paths, kinetic traps, and experimentally realizable terminal aggregates.^{30–32}

Determination of assembly pathways describing the sequence of structural steps leading to the formation of a terminal aggregate provides a powerful means for inferring the structural mechanisms underpinning assembly.²² Wilber et al.^{27,28} observed the self-assembly of patchy particles into icosahedral aggregates to proceed by two distinct assembly paths. At high temperatures, aggregates formed through stepwise addition of monomers, whereas at low temperatures, the particles flocculated into a disordered liquid phase from which near complete icosahedral clusters periodically “bud off”. The

Received: January 12, 2014

Revised: March 21, 2014

Published: March 24, 2014



authors identified these two pathways by visual inspection of simulation trajectories supported by tracking of the cluster size distribution. This approach to assembly pathway inference, however, cannot reveal the full microscopic mechanistic details of the process. A systematic approach to assembly pathway inference developed by Jankowski and Glotzer^{21,22} discovers metastable structures by computing approximate partition functions and identifies assembly pathways by constructing a directed network over these states. The great strength of this approach is that it recovers thermodynamic assembly pathways independently of system dynamics, but by the same token cannot capture the thermodynamically *and* kinetically meaningful assembly pathways within a single unified framework.

We present a new approach to infer systematically self-assembly pathways by nonlinear machine learning of molecular simulation trajectories. By mimicking the true building block dynamics within the simulations, we naturally recover assembly pathways that reflect both the thermodynamics (*what* can assemble?) and kinetics (*how* does it assemble?) of the process. As we discuss below, the nonlinear machine learning technique that we employ in this work, diffusion maps,^{38,39} circumvents the technical difficulties that plague the application of linear approaches such as principal components analysis (PCA)⁴⁰ and recovers kinetically meaningful order parameters that characterize the collective structural rearrangements driving the long-time evolution of the assembly process.³⁹ We demonstrate our approach in an application to Brownian dynamics simulations of the self-assembly of anisotropic patchy particles into polyhedral aggregates as a well-characterized test system.^{27,28,41}

Our approach is directly extensible to systems with arbitrary dynamics and arbitrary particle geometries and chemistries. We anticipate that the systematized recovery of self-assembly pathways and mechanisms will provide a deeper understanding of self-assembly processes of scientific and technological importance and ultimately help to guide rational building block design for the fabrication of materials with desirable structural and functional properties.

2. THEORETICAL METHODS

2.1. Model. We model anisotropic “patchy particles” using the model of Zhang and Glotzer⁴¹ in which each particle comprises a large sphere decorated by smaller attractive surface patches arranged in two coaxial annular rings (Figure 1). All patchy particle visualizations were generated using VMD.⁴² This model has been used previously to study the reversible self-assembly of a range of aggregates including tetrahedral, octahedral, and icosahedral clusters.⁴¹ The patchy particle is composed of three subspecies, denoted A, B, and C, that together comprise a rigid body. The larger A-spheres represent a spherical colloidal particle. The smaller B- and C-spheres that adorn the A-sphere mediate the anisotropic interparticle interactions. The B-spheres compose the ring closest to the equator of the A-sphere, while the C-spheres make up the ring closer to the pole. B–B and C–C interactions between spheres in different particles are attractive, serving as anisotropic “sticky patches”. We can alter the relative geometry of our patchy particles by varying the polar angle of the B and C rings. Following Zhang and Glotzer,⁴¹ we bias the assembly of particular polyhedra by placing the sticky patches such that they optimally contact in the geometry of an idealized aggregate. We are interested in two simple, but nontrivial, target structures upon which to test and validate our methodology: tetrahedral aggregates as a simple test case for the assembly of small

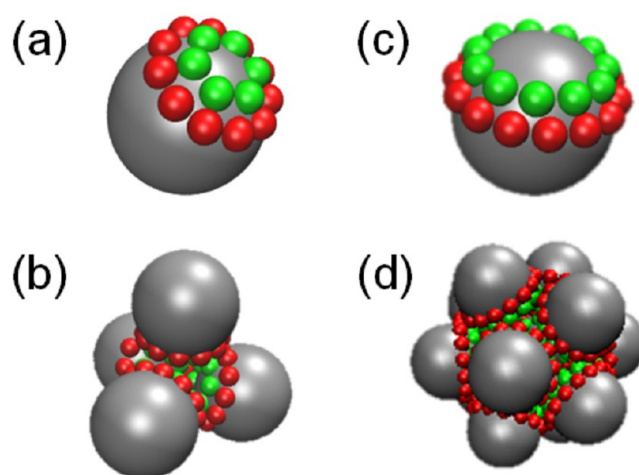


Figure 1. Rigid body double-ring patchy particles employed in this study.⁴¹ Panels (a) and (b) correspond to the structure of the tetrahedral monomer and the idealized tetrahedral cluster, respectively; (c) and (d) represent the icosahedral monomer and idealized icosahedral cluster, respectively. A-spheres are gray, B-spheres red, and C-spheres green.

polyhedral clusters and icosahedral aggregates as a toy model for viral capsid formation that has been well-characterized by Wilber et al.^{27,28} To favor icosahedral clusters, we place 14 B-spheres at a polar angle of 70.6° and 11 C-spheres at a polar angle 49.0°. To favor tetrahedral cluster formations, we place 11 B-spheres at a polar angle of 46.8° and 6 C-spheres at a polar angle of 24.9°.

We treat the patchy particles as rigid bodies neglecting all intraparticle interactions and model interparticle interactions using potentials previously implemented by Glotzer and co-workers.^{7,41} Interparticle B–B and C–C interactions are modeled by a Lennard-Jones potential

$$U_{ij}^i(r) = 4\epsilon_i \left[\left(\frac{\sigma_i}{r} \right)^{12} - \left(\frac{\sigma_i}{r} \right)^6 \right] \quad (1)$$

where σ_i is the diameter of particle i , ϵ_i the potential well depth for particle type i , and r the interparticle separation. A–X, where $X \in \{A, B, C\}$, and B–C interactions occur through a surface shifted Weeks–Chandler–Andersen (WCA) potential to account for excluded volume of the A spheres⁴³

$$U_{WCA}^{ij}(r) = \begin{cases} 4\epsilon_{ij} \left[\left(\frac{\sigma}{r - \Delta_{ij}} \right)^{12} - \left(\frac{\sigma}{r - \Delta_{ij}} \right)^6 \right] + \epsilon_{ij}, & r < 2^{(1/6)}\sigma + \Delta_{ij} \\ 0, & r \geq 2^{(1/6)}\sigma + \Delta_{ij} \end{cases} \quad (2)$$

where $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ is defined according to the Berthelot combining rule,⁴⁴ $\Delta_{ij} = (\sigma_i + \sigma_j)/2 - \sigma$,⁴⁵ and σ is a parameter. We perform our simulations in dimensionless units such that $\sigma = \sigma_B = \sigma_C = 1$, $\sigma_A = 5$, $\epsilon = \epsilon_A = \epsilon_B = \epsilon_C = 1$, and $m = m_A = m_B = m_C = 1$ where m_i is the mass of sphere i . The reduced temperature is defined as $T^* = k_B T / \epsilon$, and reduced time as $t^* = t / (m \sigma^2 / \epsilon)^{1/2}$. We note that for this choice of parameters, $\Delta_{AA} = 4$, $\Delta_{AB} = \Delta_{AC} = 2$, and $\Delta_{BC} = 0$, which effectively prohibits A–A overlaps, assuring that the patchy particles interact primarily by specific B–B, C–C, and B–C interactions on the surface of the A spheres.

2.2. Simulations. Simulations were conducted using the HOOMD-blue GPU-accelerated molecular dynamics code⁴⁶ using a rigid body Brownian dynamics integrator to implicitly model solvent effects on the motion of colloidal patchy particles in suspension.⁴⁷ The code is available at: <http://codeblue.umich.edu/hoomd-blue>. The equations of motion were numerically integrated using a symplectic, microcanonical scheme for rigid bodies developed by Miller et al.⁴⁸ in which drag and random forces are applied to each particle according to the fluctuation–dissipation relation at the specified system temperature. Periodic boundary conditions in a cubic box were implemented in all dimensions.

All simulations are initialized and performed in the following manner. First, a random initial configuration was generated by placing 120 patchy particles on a square lattice and evolving this system at high temperature ($T^* = 1.0$) with only WCA interactions active. The initial volume fraction of the box is $\phi = 0.05$, and the box dimensions are shrunk, scaling the particle positions with the box size, to reach the target volume fraction of $\phi = 0.20$ over 10^4 time steps with $dt^* = 10^{-4}$. Our choice of target volume fraction is identical to that employed by Zhang and Glotzer⁴¹ and corresponds to a particle number density of $1.8 \times 10^{-3} \sigma^{-3}$. Once the box has reached its terminal size, an additional 10^6 time steps with $dt^* = 10^{-3}$ are conducted to randomize the starting configuration. To investigate the effect of temperature on icosahedral cluster formation, we perform three independent simulations at each of nine dimensionless temperatures $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$ for 3×10^6 time steps with $dt^* = 5 \times 10^{-3}$. For temperatures $T^* \leq 0.5$, the simulations become kinetically arrested in a large disordered aggregate comprising all particles in the box. For temperatures $T^* > 0.95$, we observe a monomeric vapor of particles with transient formation of small clusters. The simulation duration was sufficiently long to observe the formation of complete icosahedra at all temperatures. Assembly of the icosahedra was effectively irreversible on the time scale of our simulations, with disassembly of complete icosahedral aggregates never observed. In the case of tetrahedral cluster formation, we perform three independent simulations at each of eight dimensionless temperatures $T^* = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ for 2×10^6 time steps with $dt^* = 5 \times 10^{-3}$. For temperatures $T^* < 0.2$, the system condenses into a disordered liquid comprising all particles in the system.

To define a mapping of our dimensionless simulations to real units, we may specify (for example) $\sigma_A = 1 \mu\text{m}$, $\rho_A = 1 \text{ g/cm}^3$, and $\varepsilon = 1$ at $T = 298 \text{ K}$. This maps our reduced temperature $T^* = 1.0$ to a real temperature of $T = 298 \text{ K}$, and the dimensionless simulation time step of $dt^* = 5 \times 10^{-3}$ to a real time step of $dt = 0.36 \mu\text{s}$. Our simulations of tetrahedral cluster assembly therefore correspond to a run time of $t = 0.7 \text{ s}$ and a temperature range of $T = 60\text{--}268 \text{ K}$, and that of icosahedral cluster assembly to $t = 1.1 \text{ s}$ and $T = 164\text{--}283 \text{ K}$.

2.3. Cluster Determination. Over the course of a simulation, snapshots of the system are saved every 10^4 time steps. To identify clusters of particles, we abstract each snapshot into interaction graphs describing the connectivity between particles in the system, where vertices and edges represent the patchy particles and the bonds between these particles, respectively.⁴⁹ A simple distance criterion is used to establish the existence of a bond in our simulation, whereby if two attractive B- or C-spheres on distinct patchy particles are within a distance of 1.5σ , the particles are considered bonded.

We have verified that our results are robust to variations in the distance threshold over the range $1.125\text{--}1.875\sigma$. We identify the strongly connected subgraphs corresponding to distinct clusters in each snapshot using Tarjan's algorithm.⁵⁰

2.4. Cluster Distance Metric. In order to apply diffusion maps to data mine assembly pathways from our simulations (see Section 2.5), we require a method to quantify the structural similarity of clusters as a proxy measure for their kinetic proximity.^{39,51} A variety of distance metrics are possible,^{52–55} but an appropriate choice must properly treat the many-body nature of the system in which clusters comprise variable numbers of identical particles. First, an N particle cluster can be described by $N!$ equivalent relabelings of the indistinguishable and interchangeable constituent particles, requiring that the measure explicitly confront this multiplicity in defining cluster similarities. Second, each cluster can have an arbitrary number of particles in an arbitrary configuration, making quantitative comparisons between clusters more complex as compared to comparisons between clusters containing the same number of particles.⁵⁶ The use of a real space Euclidean distance metric is frustrated by the absence of a spatially invariant basis with which to describe arbitrary cluster configurations containing arbitrary numbers of particles. Simple metrics comparing cluster size or degree of connectivity discard information on the internal cluster structure. More advanced metrics based on shape matching have been developed,⁵⁷ but these methods also effectively coarse grain the underlying bonding network, discarding information on the local bonding structure of the cluster in favor of cluster symmetries and archetypal structures. In principle, the earth mover's distance (EMD)⁵⁸ could be applied to define pairwise cluster similarities, but the particle “democracy” requires an exhaustive consideration of all $N!$ particle labelings, making the problem computationally intractable for large clusters.^{52,59}

We overcome the technical challenges arising from the lack of an invariant basis and particle fungibility by abstracting the cluster-matching problem into a graph-matching problem,^{22,49} where scalar measures of graph similarity between graphical representations of cluster pairs serve as good proxy measures for their structural similarity. This approach eliminates the difficulty of comparing dissimilarly sized clusters and circumvents the absence of a spatially invariant real-space basis by representing the cluster in a transformed graph space. Finding the optimal alignment of two graphs remains, however, an NP-complete problem,^{55,60} requiring the evaluation of all possible $N!$ permutations for N node graphs. While an exhaustive search is possible for small cluster sizes, this computation is intractable for large cluster graphs. For example, exhaustive comparison of two unique 12 particle clusters would require $12! = 479\,001\,600$ separate calculations to guarantee recovery of the optimal pairwise alignment. Because of the large number of unique clusters produced in our self-assembly simulations, we instead appeal to an approximate greedy solution using an adaptation of the spectral graph-matching algorithm IsoRank.⁵⁵

Our algorithm proceeds as follows. For each cluster observed in our simulations, we compute the binary adjacency matrix G_i defining the graph representing the pairwise bonding in cluster i . $G_i(p,q) = 1$ indicates that our distance criterion has defined a bond between particles p and q in the cluster, whereas $G_i(p,q) = 0$ indicates that no bond exists between these particles. We compute the global alignment between pairs of interaction graphs G_i and G_j using a slightly modified version of the IsoRank algorithm⁵⁵ that preserves local connectivities within

each cluster. The details of this modification are described in the Supporting Information. This operation defines a permutation matrix \mathbf{H} that rearranges the rows and columns of \mathbf{G}_i such that the transformation $\mathbf{G}'_i = \mathbf{H}\mathbf{G}_i\mathbf{H}^T$ represents the best global alignment of \mathbf{G}_i and \mathbf{G}_j discovered by IsoRank. Given this nodal permutation, \mathbf{H} , we transform the interparticle distance matrix between all particles in cluster i , \mathbf{R}_i , into the basis of our transformed bonding network as $\mathbf{R}'_i = \mathbf{H}\mathbf{R}_i\mathbf{H}^T$. We then define the similarity of the two clusters by summing over the absolute differences of corresponding interparticle distances between the two clusters

$$d_{ij} = \sum_{p=1}^N \sum_{q=p+1}^N |R'_i(p, q)G'_j(p, q) - R_j(p, q)G_i(p, q)| \quad (3)$$

where \mathbf{d} is the (symmetric) matrix of cluster pairwise distances, $R'_i(p, q)$ the distance between particles p and q in the transformed basis of cluster i , $G'_j(p, q)$ the corresponding element of the binary adjacency matrix in the transformed basis, $R_j(p, q)$ the distance between particles p and q in cluster j , $G_j(p, q)$ the corresponding element of the binary adjacency matrix, and $N = \max(N_i, N_j)$, where N_i is the number of particles in cluster i . Multiplying the distances by the corresponding element of the binary adjacency matrix assures that only particle pairs between which a bond has been defined participate in the sum. If $N_i < N_j$ ($N_j < N_i$), we augment \mathbf{G}'_i (\mathbf{G}_j) and \mathbf{R}'_i (\mathbf{R}_j) with $|N_i - N_j|$ additional rows and columns of zeros to represent the “missing” particles relative to the larger cluster. The product of interparticle distances to these “ghost” particles and the corresponding element of the augmented binary adjacency matrix are effectively represented within the summation as zero-length bonds, reflecting the absence of these contacts within the smaller cluster. This distance metric thus encapsulates both the local deviations in bonding as particles fluctuate around their equilibrium separation and the global differences between clusters with different bonding networks.

To efficiently compute pairwise similarities, we categorize clusters into unique geometry classes based on the cluster’s graphical representation. Two clusters belong to the same geometry class if there is a bijective mapping between vertices in the two binary adjacency matrices with zero Frobenius norm

$$\|\mathbf{G}'_i - \mathbf{G}_j\|_F = \sqrt{\sum_p \sum_q [G'_i(p, q) - G_j(p, q)]^2} \quad (4)$$

Accordingly, the bonding structure of all clusters in a geometry class is identical, although there may be deviations in bond lengths. Our implementation of IsoRank (see Supporting Information) enables rapid determination of same sized clusters belonging to the same geometry class. This categorization drastically reduces the number of pairwise alignments that must be calculated because we need to compute only the permutation matrix \mathbf{H} between each pair of unique geometry classes.

2.5. Diffusion Maps. Nonlinear machine learning has been previously applied to molecular simulation trajectories to identify the pathways by which polymers collapse and proteins fold,^{61–66} to define order parameters for the crystallization of colloidal clusters with a constant number of particles,⁵⁶ and to characterize the binding of ligands to DNA.^{67,68} In contrast to simpler linear techniques, such as PCA,^{40,69} nonlinear manifold learning techniques are basis-free, requiring only a pairwise distance measure between clusters and allowing us to

circumvent the absence of a spatially invariant basis set with which to describe many-body self-assembly phenomena.³⁹ Application of nonlinear techniques to the study of generalized self-assembly has, however, been frustrated by difficulties in defining a meaningful measure of the pairwise similarity of clusters containing differing numbers of identical particles. As described in Section 2.4, we have confronted this difficulty using graph-matching techniques, and in this work apply a nonlinear manifold learning approach known as diffusion maps^{38,39,51,54} to identify the fundamental underlying collective modes governing the long-time dynamics of patchy particle self-assembly. In analogy to the Mori–Zwanzig projection operator formalism,^{70,71} this algorithm assumes that the underlying dynamical modes for the system can be described by a (small) number of slow modes governing the long-time dynamical evolution of the system, to which the remaining fast modes are effectively slaved. Projection of the simulation trajectories into this kinetically meaningful “slow subspace” reveals the structural progression by which the system assembles into its terminal aggregates, naturally uncovering the underlying assembly pathways and mechanisms.

The application of diffusion maps to molecular simulations has been described in detail in refs 39 and 61. In brief, the diffusion map approach proceeds by generating a right-stochastic Markov matrix from our symmetric matrix of pairwise cluster similarities, \mathbf{d} . Large elements of \mathbf{d} correspond to very dissimilar clusters and small values to similar clusters. A Gaussian kernel with bandwidth ϵ is used to soft-threshold the distance matrix, effectively attenuating distances larger than $\sqrt{\epsilon}$, thereby retaining only local information on cluster proximity, to form the soft-thresholded pairwise distances matrix \mathbf{A}

$$A_{ij} = \exp\left(-\frac{d_{ij}^2}{2\epsilon}\right) \quad (5)$$

The impact of this operation is to eliminate large distances which are not expected to be dynamically meaningful.⁶¹ Because d_{ij} is a non-negative measure (cf. eq 3), the elements of \mathbf{A} are restricted to the range $0 \leq A_{ij} \leq 1$, with large values corresponding to similar clusters and small values to dissimilar clusters. We then row normalize \mathbf{A} matrix to form a Markov matrix \mathbf{M}

$$D_{ij} = \begin{cases} \sum_{k=1}^N A_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$\mathbf{M} = \mathbf{D}^{-1}\mathbf{A} \quad (7)$$

The eigenvectors of the \mathbf{M} matrix, $\{\psi_i\}_{i=1}^N$, are discrete approximations to the eigenfunctions of the Fokker–Planck operator describing a diffusion process over the data.^{38,54} Each eigenvector corresponds to a particular collective dynamical mode, with those eigenvectors with large associated eigenvalues corresponding to slow modes governing the long-time system evolution, and those with small eigenvalues to fast motions. By virtue of the Markov property, the eigenvalues are bounded in the range $[0, 1]$. We arrange these eigenvectors in nonascending order of the magnitude of their associated eigenvalue and discard the top eigenvector possessing unit eigenvalue as the trivial all-ones vector. By identifying a gap in the eigenvalue

spectrum, we isolate those collective modes above the spectral gap as driving the long-time system evolution and to which those modes below the gap are effectively slaved. In analogy with the Mori–Zwanzig formalism,^{70,71} the eigenvectors above the gap constitute the slow subspace. Geometrically, we conceive these eigenvectors as defining a low-dimensional manifold to which the dynamical evolution of the system is effectively restrained.⁶¹

The “diffusion mapping” is the embedding of the i th cluster into this low-dimensional space by projecting it into the i th component of the top k nontrivial eigenvectors

$$\mathbf{x}(i) \rightarrow (\Psi_2(i), \Psi_3(i), \dots, \Psi_{k+1}(i)) \quad (8)$$

This low-dimensional manifold describes the collective motions driving self-assembly, with each eigenvector corresponding to a particular dynamical mode. As described in the Supporting Information, the systems in this work each exhibited a gap in the eigenvalue spectrum at either $k = 2$ or 3 (Figures S3 and S4 of Supporting Information).

As discussed in ref 61, the leading eigenvectors of the \mathbf{M} matrix correspond to the slow collective modes of the system under two relatively mild assumptions: (i) the system dynamics are well-modeled by a diffusion process and (ii) that our pairwise similarity measure between clusters is a good measure of short-time diffusive motions. A priori, we anticipate that interactions between the patchy colloids will lead to cooperative couplings between the system degrees of freedom and the emergence of a small number of slow collective modes governing the long-time self-assembly behavior. The presence of a gap in the eigenvalue spectrum of the \mathbf{M} matrix provides post hoc substantiation of this conjecture. This separation of time scales suggests that the system dynamics may be modeled under the Mori–Zwanzig formalism as a set of coupled stochastic differential equations in the slow collective modes, to which the remaining fast degrees of freedom couple as stochastic noise. Such a description suggests that the system may be modeled as a diffusive process, satisfying the first assumption.

Our measure of pairwise cluster similarities by graph matching captures both global differences resulting from addition or removal of particles from the cluster and local differences stemming from bond length fluctuations, furnishing a measure of the “structural remodeling” required to transform one cluster into another. Small values of this measure are therefore expected to reflect short-time changes in the cluster structure resulting from fluctuations in interparticle distances and/or number of particles in the cluster. Accordingly, we expect our cluster similarity metric to satisfy the second assumption by serving as a good measure of short-time diffusive evolution of the cluster structure.

Under these two assumptions, the leading eigenvectors correspond not only to the slow collective modes but also to the Euclidean distances between clusters in the embedding prescribed by eq 8 approximate “diffusion distances” in real space, where the diffusion distance can be interpreted as the ease with which one cluster may transform into another.^{38,61,51} (When all $(N - 1)$ nontrivial eigenvectors in the embedding are included, this correspondence becomes exact.⁵¹) Together, these two properties imbue the diffusion map with its capacity to project simulation trajectories into dynamically meaningful low-dimensional embeddings that map out the long-time evolution of the system and, as a corollary, reveal the underlying self-assembly pathways and mechanisms. A number

of nonlinear manifold learning techniques exist, including Isomap,⁷² local linear embedding (LLE),⁷³ and Laplacian eigenmaps,⁷⁴ but it is this attractive feature of diffusion maps that lead us to favor this approach.

2.6. Physical Order Parameters. Despite these very attractive properties of diffusion map embeddings, it is a well-known deficiency of diffusion maps (and of nonlinear dimensionality reduction methodologies in general) that the approach does not also provide a direct physical interpretation of the slow collective modes approximated by the leading eigenvectors of the \mathbf{M} matrix.³⁹ Thus, while order parameters furnished by the diffusion map are good descriptors of the slow collective modes, assigning physical meaning to these modes is a nontrivial task.^{61,62} In this work, we derive physical interpretability by correlating each eigenvector with candidate physical variables to furnish physical “bridge” variables to help physically interpret the diffusion map embeddings.⁶¹

As described below, we identified the Steinhardt bond orientation order parameters $\{Q_l\}$ ⁷⁵ as good bridge variables which have been used extensively in the analysis of Lennard–Jones cluster formations⁷⁶ and in nucleation theory.^{77,78} Q_l represents the l th order sum of rotationally invariant spherical harmonics and is given by

$$Q_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{Q}_{lm}|^2 \right]^{1/2} \quad (9)$$

where \bar{Q}_{lm} is defined as

$$\bar{Q}_{lm} = \frac{1}{N_b} \sum_{\vec{r}_b} Y_{lm}(\theta(\vec{r}_b), \phi(\vec{r}_b)) \quad (10)$$

where \vec{r}_b is the vector connecting the cluster center of mass to a bond midpoint and θ and ϕ are the polar and azimuthal angles corresponding to this vector \vec{r}_b ; Y_{lm} is the spherical harmonic of degree l and order m , and N_b is the number of bonds in the cluster, and the sum is over \vec{r}_b runs over all N_b vectors. By projecting cluster orientations onto rotationally invariant combinations of spherical harmonics these order parameters effectively measure the degree of angular ordering of the particles around the cluster center of mass.

In Section 3.3 we discuss the effectiveness of these order parameters in characterizing the coordinates furnished by the diffusion map for the self-assembly of icosahedral clusters. It is conceivable that other physical variables exist that might show improved correlation with the diffusion map coordinates, for example, structural order parameters corresponding to the bulk crystal phases, and we observe that systematic procedures exist to screen pools of candidate variables to identify such physical variables.^{79,80} We note, however, that while the identification of such correspondences can provide a wealth of physical insight into the self-assembly process, the inherently many-body and often complex nature of the self-assembly process means that we may not expect simple physical interpretations of the underlying collective modes identified by the diffusion map.

3. RESULTS AND DISCUSSION

3.1. Tetrahedral Cluster Formation. Following the protocols described in Theoretical Methods, we simulated the self-assembly of our tetrahedron-forming patchy particles at temperatures $T^* = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ from which we identified a total of 63 335 clusters belonging to 427 unique geometry classes. For clusters belonging to the same

geometry class, the mean distance between them is 2.87 ± 0.02 , while for clusters belonging to different classes, the mean minimum distance is 7.37 ± 0.11 . This shows that our distance metric is able to effectively capture bonding differences both within and between the different cluster geometries. Differences between clusters within the same geometry class correspond to fluctuations around the equilibrium bond distances, while distances between clusters in different geometry classes also arise from the addition or removal of bonds within the cluster.

The structure of each cluster observed in our simulation is defined in real space by the (relative) position and orientation of all particles in the cluster. The low-dimensional embedding constructed by the diffusion map projects these configurations onto a manifold in which the various clusters are separated according to their structural similarity. Simulations at different temperatures are expected to explore different regions of configurational phase space because high-temperature systems can access higher-energy configurations and jump energy barriers that low-temperature systems cannot. (Indeed, because of finite sampling and sensitive dependence on initial conditions, simulations at the same temperature initialized with distinct starting configurations are also expected to explore different configurational ensembles.)

Because the manifold constructed by the diffusion map is a function of the configurations from which it is constructed (cf. eqs 5–8), to compare the low-dimensional embeddings of different simulations in the same basis set, it is necessary to construct a single composite manifold from the configurations aggregated from all simulation trajectories. Constructing a different manifold for each simulation trajectory would prohibit comparisons across simulations because the basis sets (i.e., the eigenvectors Ψ_i spanning each embedding) are constructed from different structural ensembles and are therefore not transferable between embeddings. The various regions of the composite manifold explored by different simulations can then be identified by considering only those data points corresponding to those runs. In this manner, the regions of the composite manifold explored by simulations at high and low temperatures can be identified. Because the volume of thermally accessible phase space increases with temperature, we anticipate that at equilibrium the range of structures explored by low-temperature simulations will constitute a subset of those sampled at high temperatures. As a corollary, we expect the volume of the composite manifold sampled by a simulation at low temperature to be a subvolume of that explored at a high temperature.

We applied the diffusion map to the complete ensemble of 63 335 clusters identified at all eight temperatures to construct a single composite diffusion map embedding of the self-assembly process. The soft-thresholding bandwidth was selected using the approach of Coifman et al.⁸¹ (Figure S1 of Supporting Information). Using the L -method,⁸² we identified a gap in the diffusion map eigenvalue spectrum after the third nontrivial eigenvalue (Figure S3 of Supporting Information), motivating the construction of three-dimensional embeddings. Upon embedding into the three nontrivial eigenvectors with eigenvalues above the spectral gap, a small number of 59 clusters were identified as outliers separated from the rest of the data in the low-dimensional projection (data not shown). From the correspondence of real space diffusion distances with diffusion map Euclidean distances, these clusters may be interpreted as rarely visited configurations to which no well-sampled dynamic pathway exists to the bulk of the data, possibly because of the presence of high free-energy barriers.

Following the “deislanding” procedure detailed in ref 62, we removed these outlying clusters from our cluster ensemble and recomputed the diffusion map embedding over the remaining 63 276 clusters.

We present our diffusion map embedding in Figure 2. Each point corresponds to a single cluster colored according to the

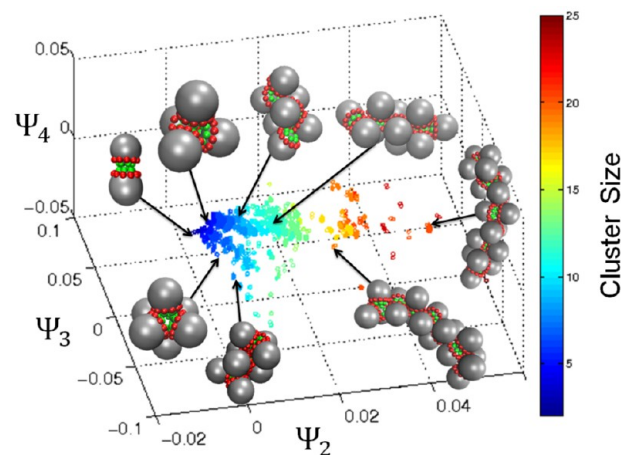


Figure 2. Three-dimensional embedding of the tetrahedral self-assembly process in the top three eigenvectors, $[\Psi_2, \Psi_3, \Psi_4]$, furnished by the diffusion map. Each point corresponds to one of 63 276 clusters observed in Brownian dynamics simulations at one of eight temperatures, $T^* = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Particular clusters have been selected for visualization to illustrate the structural evolution along the assembly pathways. Points have been colored according to cluster size.

number of particles it contains. Visualizations of representative cluster geometries along the manifold have been overlaid to reveal the assembly pathways resolved by this low-dimensional embedding. Our approach systematically infers the progression of cluster arrangements that lead from free monomers to self-assembled tetrahedral aggregates. Further, we can see the formation of diverse structures from the desired tetrahedral formation, such as hexahedra, and two divergent assembly pathways leading to the formation of two morphologically distinct particle chains. The chain visible to the right of the image corresponds to a stack of interlocking dimers and tetrahedral tetramers, while that toward the bottom is composed of a stack of interlocking trigonal planar trimers. The chain length is limited only by the number of particles present in the simulation box.

To more easily visualize the assembly pathways revealed by the diffusion map, in Figure 3 we have projected our three-dimensional embedding into the two-dimensional planes parametrized by $[\Psi_2, \Psi_3]$ and $[\Psi_2, \Psi_4]$. By collapsing the three-dimensional embedding onto the $[\Psi_2, \Psi_3]$ plane (Figure 3a), we can discern some variation in the assembly pathways leading to the two chains, but the variability with respect to Ψ_3 is insufficient to clearly separate the two assembly pathways. In contrast, the two distinct routes can be clearly discerned in the embedding in $[\Psi_2, \Psi_4]$ in Figure 3b. The top pathway corresponds to the formation of the stacked interlocking dimers and tetrahedral tetramers, while the bottom pathway yields the stacked interlocking trigonal planar trimers. Visualization of representative clusters along the pathway extracted by the diffusion map reveals the assembly mechanism along each path. In the upper pathway, monomers first condense into

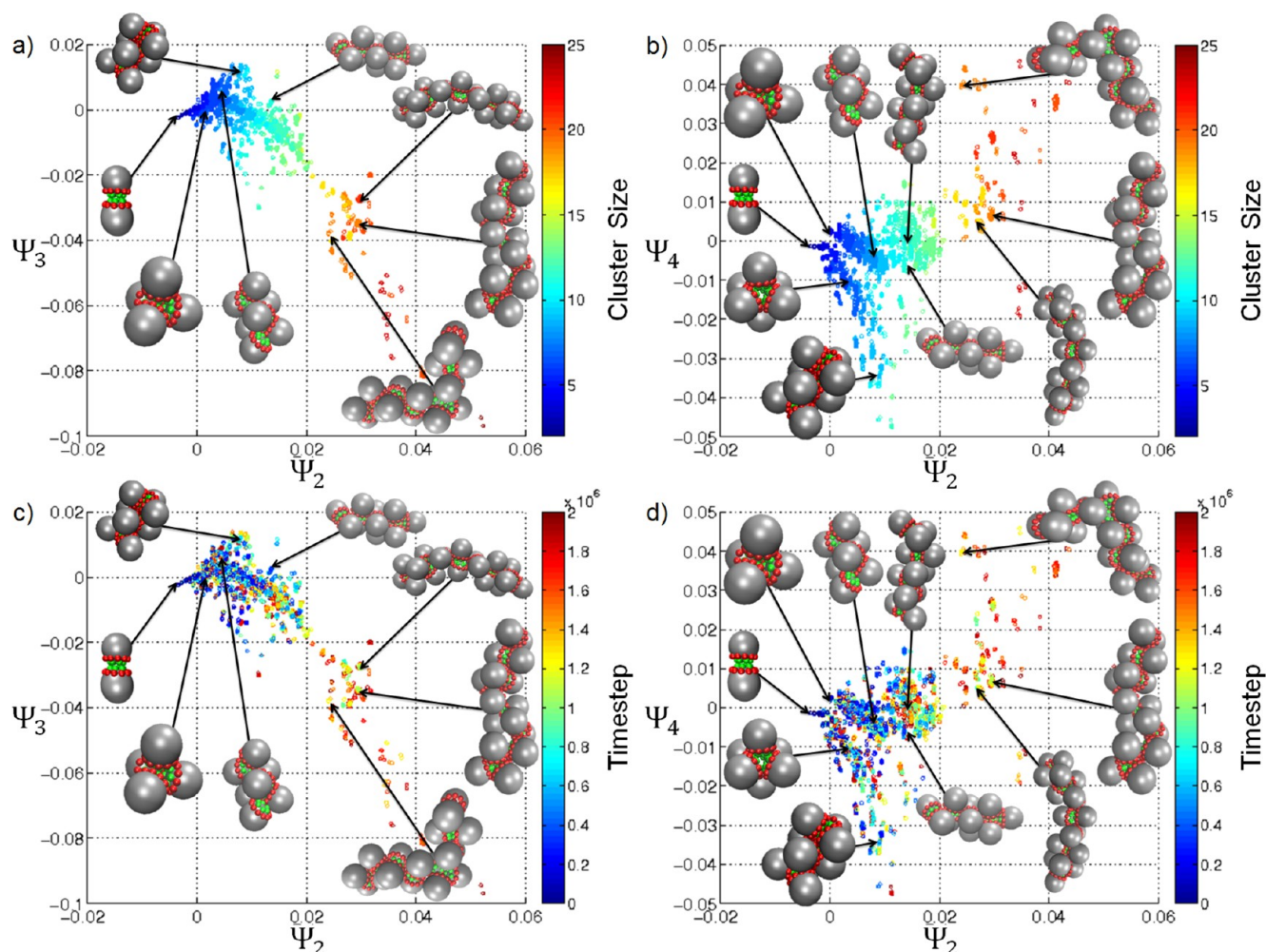


Figure 3. Two-dimensional embedding of the tetrahedral self-assembly process into (a, c) $[\Psi_2, \Psi_3]$ and (b, d) $[\Psi_2, \Psi_4]$. Each point corresponds to one of 63 276 clusters observed in Brownian dynamics simulations at one of eight temperatures, $T^* = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Particular clusters have been selected for visualization to illustrate the structural evolution along the assembly pathways. Points in panels (a) and (b) have been colored according to cluster size, and those in panels (c) and (d) according to the simulation time at which the cluster was observed.

dimers and tetrahedral tetramers, which then form dimer–dimer and tetramer–dimer pairs that condense monomers and dimers from the bulk into an extending chain. This appears to enable structural rearrangement into relative particle positions that maximize the overlap of attractive patches. The bottom pathway, on the other hand, proceeds exclusively by the condensation of free trimers and yields a more dense interlocking structure. Although the patches on the particle surfaces were placed to favor assembly of the idealized tetrahedral cluster, we see that (depending on temperature, see Figure 4) this may represent only an intermediate, rather than terminal, aggregation state due to these chain condensation processes.

In Figure 4a, we reproduce Figure 3b, and in panels b–i we restrict the embeddings to only those points harvested at each of the eight temperatures considered. By modulating the temperature, or equivalently modifying the interaction strength, we observe that we can selectively activate or deactivate the pathways for formation of chain-like aggregates. At high temperature (low interaction strength), $0.7 \leq T^* \leq 0.9$, interparticle interactions are sufficient to stabilize clusters containing up to only eight particles. The system exists as an ensemble of small dispersed clusters, with tetrahedral clusters

dominating. Upon lowering the temperature (increasing the interaction strength) to $0.5 \leq T^* \leq 0.6$, the bottom assembly pathway to chains of stacked trigonal planar trimers becomes active. In this temperature regime, dimers and trimers readily form. Collision of two trimers induces a structural relaxation that leads the trimers to interlock, permitting partial overlap of the attractive patches and energetic stabilization of the aggregate. Further elongation of the interlocking chain can occur through the addition of free trimers, as well as monomers and dimers. Coalescence of monomers at the chain end caps the chain by forming an effective tetrahedral aggregate at the terminus. Similarly, interaction of the chain terminus with dimers can induce the dimer to decompose, with one partner serving as the end-cap while the other is released back into the bulk. Capped trimer chains are able to continue growing through the addition of a monomer that binds to the tetrahedral terminus, serving as a bridging site for further aggregation.

Upon lowering the temperature further (increasing the interaction strength), $0.2 \leq T^* \leq 0.4$, assembly of chains of interlocking dimers and tetrahedral tetramers becomes favorable and begins to compete with the assembly of trimer chains. In this regime, bonding between monomers is

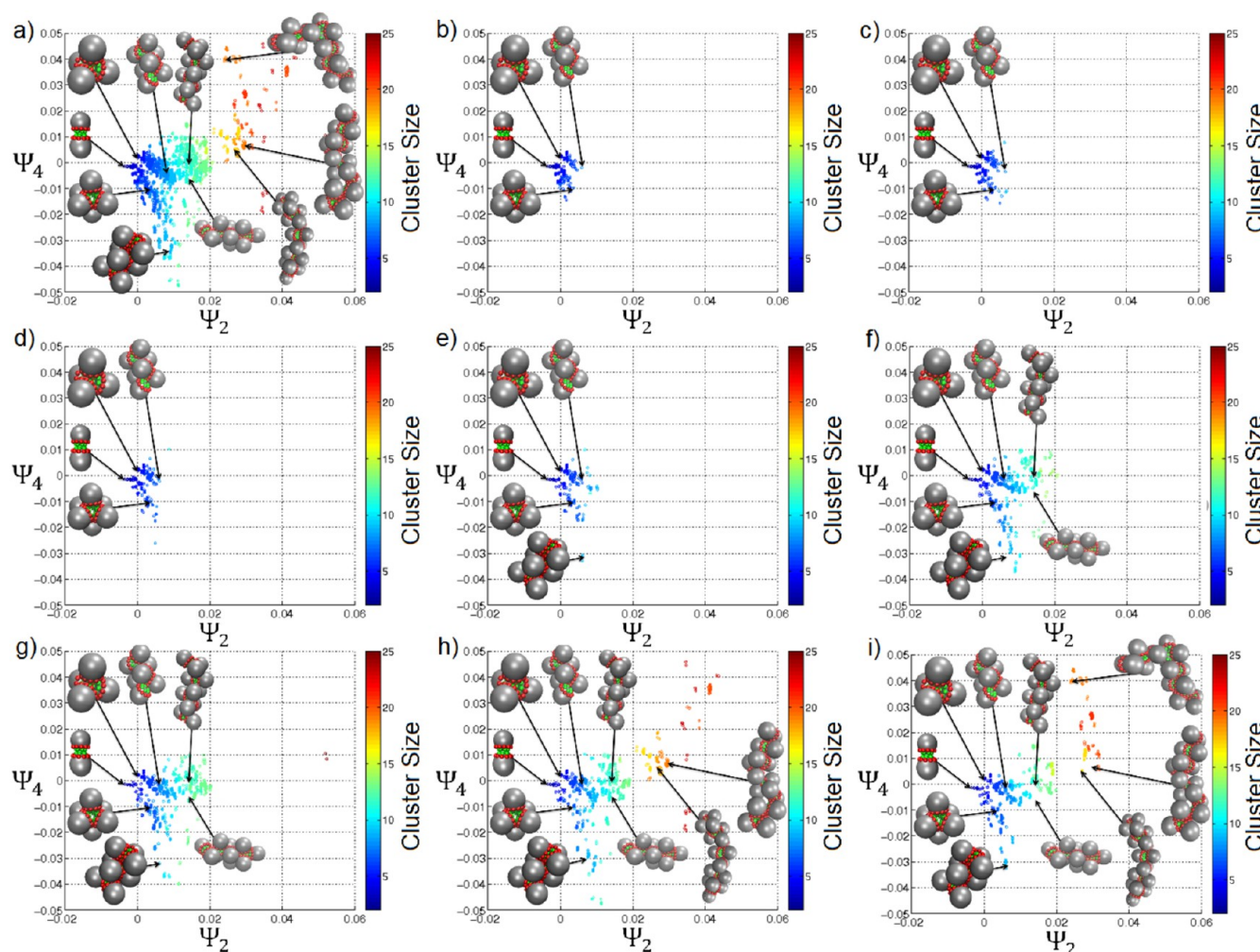


Figure 4. Two-dimensional embedding of the tetrahedral self-assembly process into $[\Psi_2, \Psi_4]$ at (a) all eight temperatures $T^* = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, (b) $T^* = 0.9$, (c) $T^* = 0.8$, (d) $T^* = 0.7$, (e) $T^* = 0.6$, (f) $T^* = 0.5$, (g) $T^* = 0.4$, (h) $T^* = 0.3$, and (i) $T^* = 0.2$. Points are colored according to cluster size, and particular clusters are selected for visualization.

effectively irreversible and dimers quickly form from free monomers. The collision of a dimer with another dimer or larger aggregate leads to condensation, where the dimer reorients to an interlocking position that maximizes attractive patch overlap in the aggregate. The formation of tetrahedra can still occur in this low-temperature setting, but they are readily condensed into the interlocking chain structures.

By embedding our simulation trajectories using diffusion maps, we have systematically identified two distinct assembly pathways and mechanisms leading to the formation of long chains of particles. Although we designed the anisotropic interactions of our patchy particle to favor the assembly of idealized tetrahedral clusters, we observe formation of tetrahedral aggregates as only one component of a mixture of small ($N < 8$) clusters at high temperatures (weak interaction strengths), $0.7 \leq T^* \leq 0.9$, or in addition to long chains of stacked dimers and tetrahedral clusters at lower temperatures (stronger interaction strengths), $0.2 \leq T^* \leq 0.4$. At intermediate temperatures, $0.5 \leq T^* \leq 0.6$, tetrahedral clusters appear metastable relative to long chains of interlocking trigonal planar trimers. These findings demonstrate the rich range of cluster morphologies and assembly mechanisms that can be accessed by manipulating the interparticle interaction strength. However, modification of interaction strength alone is

insufficient to uniquely favor the assembly of tetrahedral clusters over all competing aggregates, indicating that particle redesign by adjusting placement of the anisotropic interaction patches will be necessary to achieve this goal. We anticipate that placing the annular rings closer to the particle poles will better mask the patches within the tetrahedral cluster, discouraging the formation of extended chains. Placing the rings too close to the pole, however, will likely favor dimers over tetramers.

We note that the simulations conducted in this work, particularly at the lower temperatures considered, are intrinsically nonequilibrium, in which we simulate the irreversible assembly of large aggregates. Accordingly, while the assembly pathways unveiled by the diffusion map reveal the underlying kinetic mechanisms by which self-assembly proceeds, the simulation snapshots are not drawn from an equilibrium distribution; therefore, each cluster is not represented in proportion to its Boltzmann factor. As a result, the density of points in the diffusion map embeddings cannot be related to the underlying equilibrium probability distribution, and it is not possible to infer the relative prevalences of the observed clusters and/or compute empirical free-energy surfaces over the embeddings.^{39,61} The assembly pathways revealed in the present work may be viewed as the progression of the self-assembly process along the low-lying valleys of the underlying

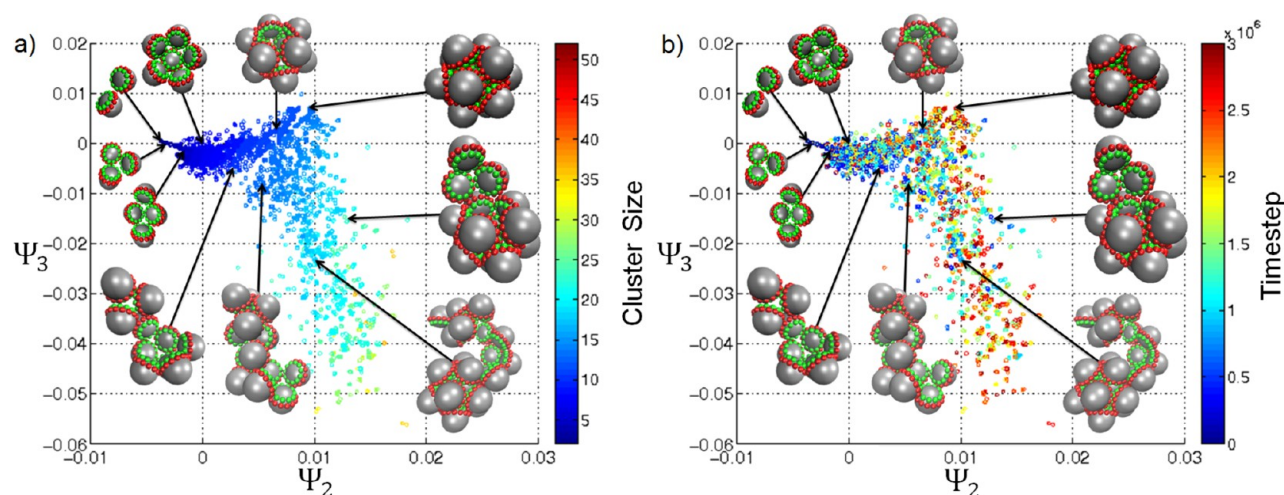


Figure 5. Two-dimensional embedding of the icosahedral self-assembly process in the top two eigenvectors, $[\Psi_2, \Psi_3]$, furnished by the diffusion map. Each point corresponds to one of 48 010 clusters observed in Brownian dynamics simulations at one of nine temperatures, $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$. Particular clusters have been selected for visualization to illustrate the structural evolution along the assembly pathways. Points in panel (a) have been colored according to cluster size, and those in panel (b) according to the simulation time step at which the cluster was observed.

free-energy surface. In future work we shall conduct biased sampling directly in the order parameters furnished by the diffusion map (i.e., $[\Psi_2, \Psi_3, \Psi_4]$) to recover the global topography of the free-energy landscape parametrized by these kinetically meaningful variables.^{83,84} These landscapes will reveal the relative thermodynamic stabilities of the clusters observed in our simulations, and, by taking empirical temperature derivatives,⁸⁵ we shall decompose the energetic and entropic contributions to the thermodynamic driving forces for self-assembly. Together with the kinetically meaningful assembly pathways, we expect this analysis to help rationalize the relative preference for finite-sized aggregates, dimer/tetramer chains, and trimer chains as a function of temperature.

3.2. Icosahedral Cluster Formation. **3.2.1. Identification of Different Assembly Mechanisms.** Following the same protocol as above, we conducted simulations of our icosahedron-forming patchy particles at each of nine different temperatures $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$. We identify from these simulations a total of 48 010 clusters belonging to 1365 unique geometry classes. The mean distance between clusters in the same geometry class is 3.06 ± 0.03 , whereas the mean minimum distance between different geometries is 10.08 ± 0.24 , again showing that our distance metric is able to effectively resolve in-class and out-of-class distances. The diffusion map was applied to the pairwise similarities between all 48 010 structures to produce low-dimensional embeddings of the icosahedral self-assembly process. The soft-thresholding bandwidth was selected using the approach of Coifman et al.⁸¹ (Figure S2 of Supporting Information), and a gap in the eigenvalue spectrum after the second nontrivial eigenvalue (Figure S4 of Supporting Information) motivated us to construct the diffusion map embeddings into the top two nontrivial eigenvectors presented in Figure 5.

The diffusion map embedding reveals two distinct assembly pathways leading to the same terminal icosahedral aggregates. This is qualitatively different from the tetrahedral system, in which we inferred two divergent assembly pathways leading to two morphologically distinct chains. Furthermore, the length of the extended chains of tetrahedral building blocks was limited

only by the number of particles in the simulation, whereas the icosahedral building blocks produce self-limiting, finite-sized icosahedral aggregates.

We demonstrate the impact of temperature upon assembly by extracting points from the composite embedding corresponding to only those clusters observed at each of the nine temperatures and color points according to cluster size (Figure 6), mean number of bonds per particle in the cluster (Figure 7), and simulation time step (Figure 8). At high temperatures, $T^* = 0.95$ (Figure 6b), there is no pathway from dimers to icosahedral structures as the thermal energy is on the order of the interaction strength between patchy particles, resulting in a gas of monomers, dimers, and other small aggregates.

As the temperature decreases (interaction strength increases), $0.85 \leq T^* \leq 0.90$ (Figure 6c,d), a path emerges whereby dimers can transform into icosahedra by monomeric addition in which the mean number of bonds per particle increases toward the five nearest neighbors of each particle in the terminal icosahedral cluster (Figure 7c,d). Thus, this high-temperature assembly mechanism proceeds in a stepwise fashion with free monomers accreting onto a partially formed aggregate in the structural configuration of a partially formed icosahedral cluster. This high-temperature mechanism is identical to that previously reported by Wilber et al.^{27,28} by visual inspection of assembly trajectories supported by tracking of particle size distributions.

As the temperature is lowered to $T^* = 0.80$ (Figure 6e), we observe the formation of large ($N > 12$) disordered aggregates that do not lie on the monomeric addition assembly pathway. At still lower temperatures (stronger interaction strengths), $0.65 \leq T^* \leq 0.75$ (Figure 6f–h), this ensemble of liquid-like aggregates expands, opening up a second assembly pathway to the terminal icosahedra, wherein complete icosahedral, or near-icosahedral, clusters periodically detach from the disordered liquid phase. This second assembly mechanism becomes more prevalent as the temperature decreases, lowering the fraction of clusters that assemble by monomeric addition (cf. Figure 10), eventually dominating the monomeric pathway at low temperatures, $0.55 \leq T^* \leq 0.60$ (Figure 6i,j). This low-temperature

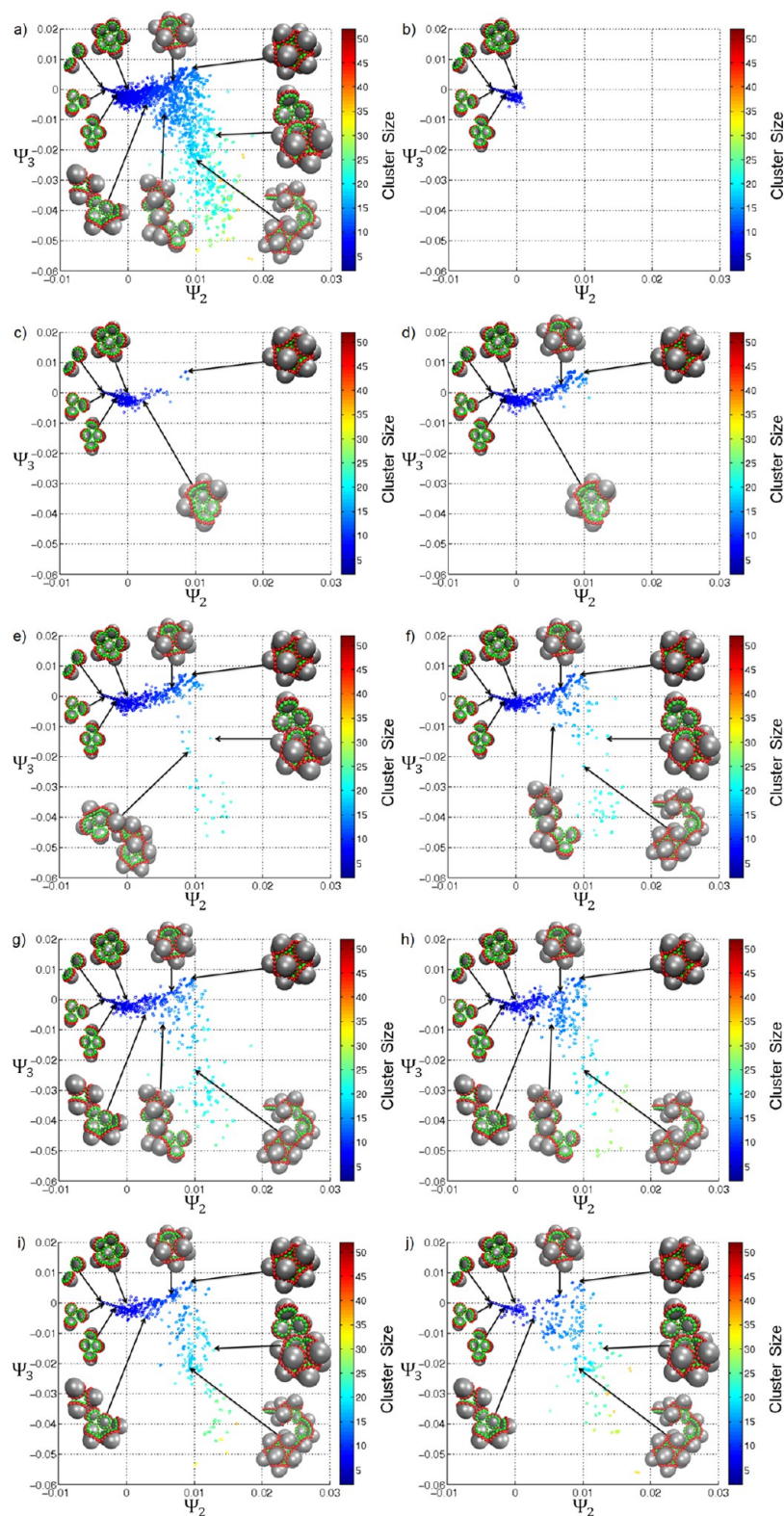


Figure 6. Two-dimensional embedding of the icosahedral self-assembly process into $[\Psi_2, \Psi_3]$ at (a) all nine temperatures $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$, (b) $T^* = 0.95$, (c) $T^* = 0.90$, (d) $T^* = 0.85$, (e) $T^* = 0.80$, (f) $T^* = 0.75$, (g) $T^* = 0.70$, (h) $T^* = 0.65$, (i) $T^* = 0.60$, and (j) $T^* = 0.55$. Points are colored according to cluster size, and particular clusters are selected for visualization.

icosahedral assembly mechanism is precisely the “budding” mechanism previously described by Wilber et al.^{27,28}

The mechanistic difference between the budding and monomeric addition pathways is clearly illustrated in Figure 7g–j. Along the upper monomeric addition route, the average

number of bonds per particle increases monotonically toward the five nearest neighbors of the terminal icosahedron, whereas along the lower budding pathway, the particles assemble into disordered clusters within which internal structural rearrangements periodically yield subclusters possessing ~ 5 bonds per

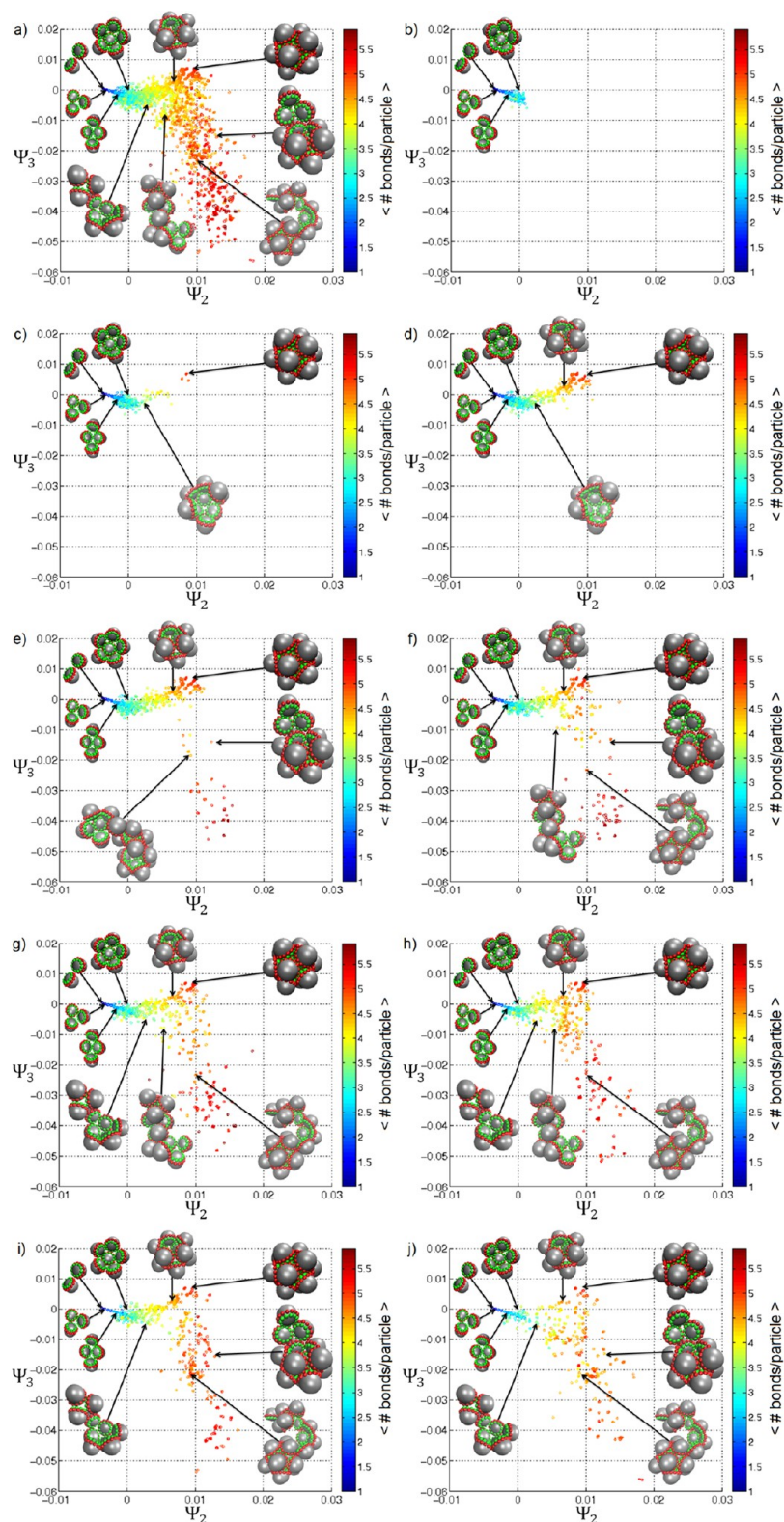


Figure 7. Two-dimensional embedding of the icosahedral self-assembly process into $[\Psi_2, \Psi_3]$ at (a) all nine temperatures $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$, (b) $T^* = 0.95$, (c) $T^* = 0.90$, (d) $T^* = 0.85$, (e) $T^* = 0.80$, (f) $T^* = 0.75$, (g) $T^* = 0.70$, (h) $T^* = 0.65$, (i) $T^* = 0.60$, and (j) $T^* = 0.55$. Points are colored according to the mean number of bonds per particle in the cluster, and particular clusters are selected for visualization.

particle that subsequently detach into the terminal icosahedral aggregates.

We have employed diffusion maps to systematically reveal and resolve two distinct assembly pathways, monomeric

addition and budding from a disordered liquid, for the assembly of terminal icosahedral clusters. Unlike the tetrahedral system, the idealized icosahedral aggregate that inspired placement of our anisotropic interaction patches does represent

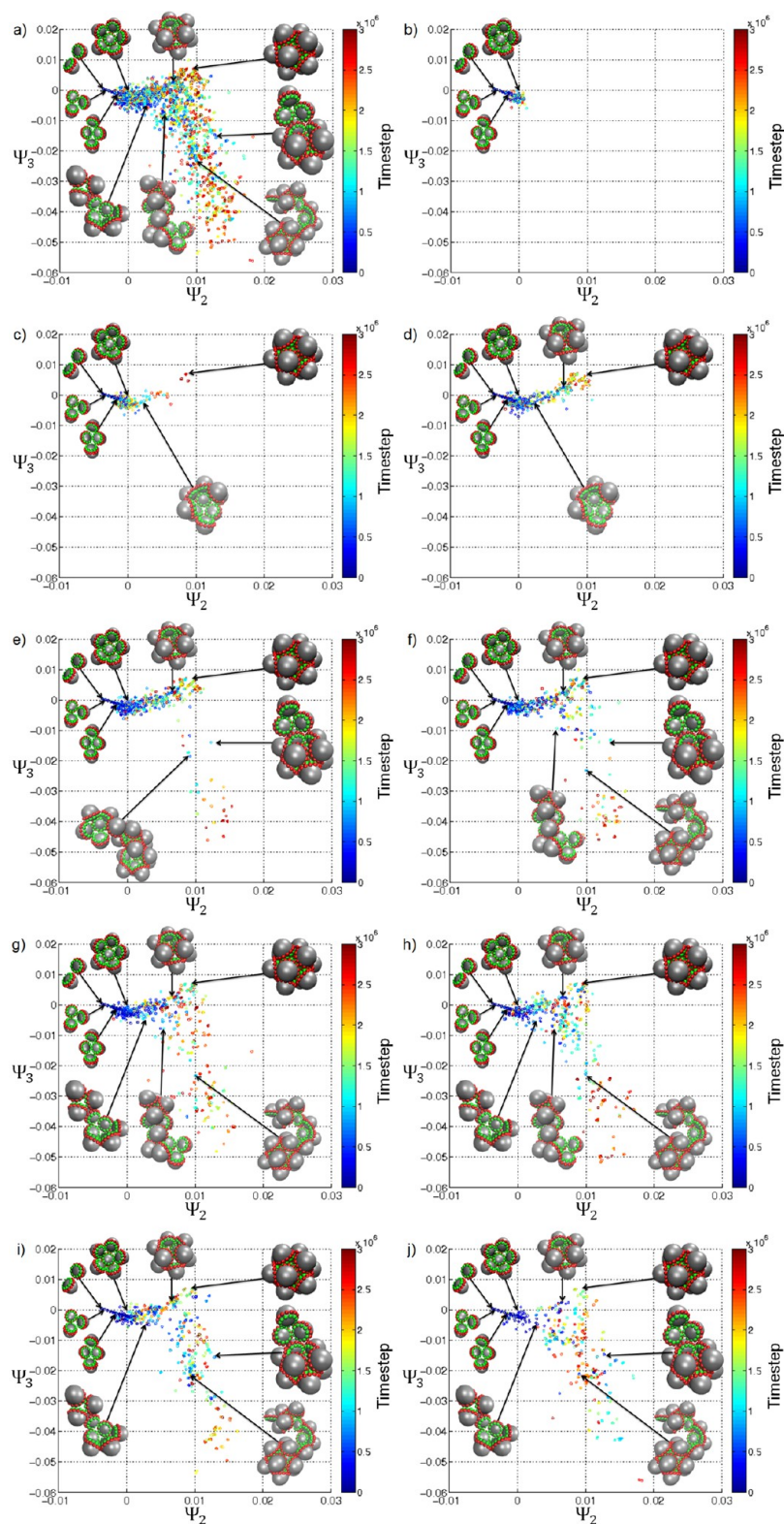


Figure 8. Two-dimensional embedding of the icosahedral self-assembly process into $[\Psi_2, \Psi_3]$ at (a) all nine temperatures $T^* = \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$, (b) $T^* = 0.95$, (c) $T^* = 0.90$, (d) $T^* = 0.85$, (e) $T^* = 0.80$, (f) $T^* = 0.75$, (g) $T^* = 0.70$, (h) $T^* = 0.65$, (i) $T^* = 0.60$, and (j) $T^* = 0.55$. Points are colored according to the simulation time at which the cluster was observed, and particular clusters are selected for visualization.

the low-temperature stable terminal aggregate at temperatures $T^* > 0.5$. At lower temperatures, the system collapses into a disordered liquid that is kinetically trapped on the time scale of our simulations. Furthermore, by modulating temperature

(interaction strength) we can control the mechanism by which the icosahedra assemble: monomeric addition at high temperatures and budding at low temperatures (Figure 10). The two competing assembly pathways identified by our

approach are in good agreement with those previously reported by Wilber et al.,^{27,28} providing strong validation that our methodology can perform automated extraction of assembly pathways and mechanisms that were previously deducible only by visual inspection of assembly trajectories.

As for tetrahedral assembly, in future work we shall recover the equilibrium free-energy surfaces by performing biased sampling in the diffusion map order parameters^{83,84} to reveal the relative thermodynamic driving forces, relative cluster stabilities, and free-energy barriers along each of the two competing assembly routes as a function of temperature.

3.2.2. Pathway Propensity. The embeddings in Figures 6–8 reveal two distinct assembly pathways but do not inform the relative assembly fluxes through each pathway at each temperature. Because assembled icosahedra are never observed to disassemble, the icosahedral state represents a sink state, and by tracing the assembly of each icosahedron backward in time, we can infer the pathway along which it passed in the low-dimensional assembly landscape and therefore by which of the two mechanisms it assembled. Because our simulations mimic the true system dynamics, this calculation presents an empirical estimate of the relative assembly fluxes along each path as a function of temperature.

To quantify the degree to which each icosahedron assembles by monomeric addition or liquid budding, we first must classify each cluster in our embedding as belonging to one of these two pathways. We use a support vector machine (SVM) classifier with a radial basis function kernel to compute a nonlinear surface, the separatrix, separating the two classes in the two-dimensional embedding in $[\Psi_2, \Psi_3]$ (Figure 5a). We construct this surface by first generating a uniform 100×100 square grid of points over the two-dimensional diffusion map embedding spanning $\Psi_2 = [-0.004, 0.023]$ and $\Psi_3 = [-0.056, 0.009]$. Each point in the grid is classified as belonging to the monomeric assembly or liquid budding mechanism by majority vote of a distance-weighted five nearest neighbor classifier comprising training points drawn from a high-temperature simulation at $T^* = 0.90$ in which we define all points to belong to the monomeric addition pathway (Figure 6c) and a low-temperature simulation at $T^* = 0.55$ in which we define all points to belong to the budding pathway (Figure 6j). We then compute the separatrix by training a two-class SVM classifier over the classified 100×100 grid using a radial basis function kernel with Gaussian bandwidth of 1.0 and box constraint for the soft margin of 1.0. We elect to apply the SVM to the classified grid trained on the high- and low-temperature data, rather than to the raw data itself, to smooth out the large deviations in density of the point cloud in our embedding (Figure 9) and synthesize a more robust and smooth classification surface. Our attempts to train an SVM on the raw data produced a classifier that was unable to effectively distinguish the two assembly mechanisms because it was disproportionately biased to generate a classification surface within the high-density region of points localized within $(\Psi_2 \approx [-0.005, 0.005], \Psi_3 \approx [-0.008, 0.002])$, effectively ignoring the sparsely populated region $(\Psi_2 \approx [0.005, 0.017], \Psi_3 \approx [-0.05, -0.01])$. Finally, we use this separatrix to classify all of the clusters observed at all temperatures by specifying that those residing below the separatrix belong to the budding path and those above belong to monomeric addition (Figure 9). Our classifier maintains the general demarcation between the two pathways described in the previous section and enables computation of the relative flux through each

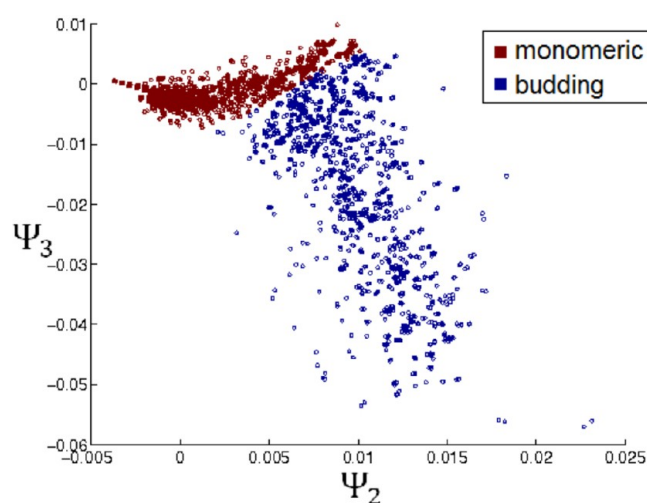


Figure 9. SVM classification of the $[\Psi_2, \Psi_3]$ diffusion map embeddings of the 48 010 clusters harvested at all nine temperatures as belonging to the monomeric addition (red) or budding (blue) assembly pathway.

pathway by unambiguously assigning each cluster to one of the two assembly pathways.

For each temperature trajectory, we identify each terminal icosahedron and trace the constituent particles back through all intermediate clusters leading to assembly. For each trace we define the fraction of budding character of the corresponding icosahedron as

$$f_{\text{budding}} = \frac{\sum_{i=1}^{N_{\text{trace}}} \text{size}(i) \text{class}(i)}{\sum_{i=1}^{N_{\text{trace}}} \text{size}(i)} \quad (11)$$

and the fraction of monomeric addition character as

$$f_{\text{monomer}} = 1 - f_{\text{budding}} \quad (12)$$

where $\text{size}(i)$ is the number of particles in the i th cluster in the trace; $\text{class}(i)$ is 1 if the i th cluster belongs to the budding class and 0 if it belongs to the monomeric addition class, and N_{trace} is the number of clusters constituting the trace that ultimately coalesces into the terminal icosahedral aggregate. For icosahedra that form predominantly through budding, f_{budding} will be close to 1, whereas for those that form principally through monomeric addition it will be close to 0.

In Figure 10a we present the mean and standard deviation of the budding and monomeric addition character averaged over all icosahedra formed at each temperature. At high temperatures, $T^* > 0.65$, the monomeric addition pathway dominates the budding pathway, whereas at low temperatures, $T^* < 0.65$, this trend inverts. At $T^* \approx 0.65$, both pathways are equally active. The monomeric addition pathway does not completely shut off in the low-temperature limit, however, as the formation of large liquid aggregates initially proceeds by the formation of small clusters by monomeric addition that subsequently coalesce into the large unstructured aggregates.

In Figure 10b we show the number of icosahedra observed to assemble in each of the three independent simulations at each temperature. From this plot it is apparent that the temperature of maximum flux is $T_{\text{max}}^* \approx 0.75\text{--}0.80$. Interestingly, in this temperature window both assembly pathways are active, with the monomeric route more populated than the budding route. The existence of a temperature optimum can be understood as a complex interplay of assembly thermodynamics (high

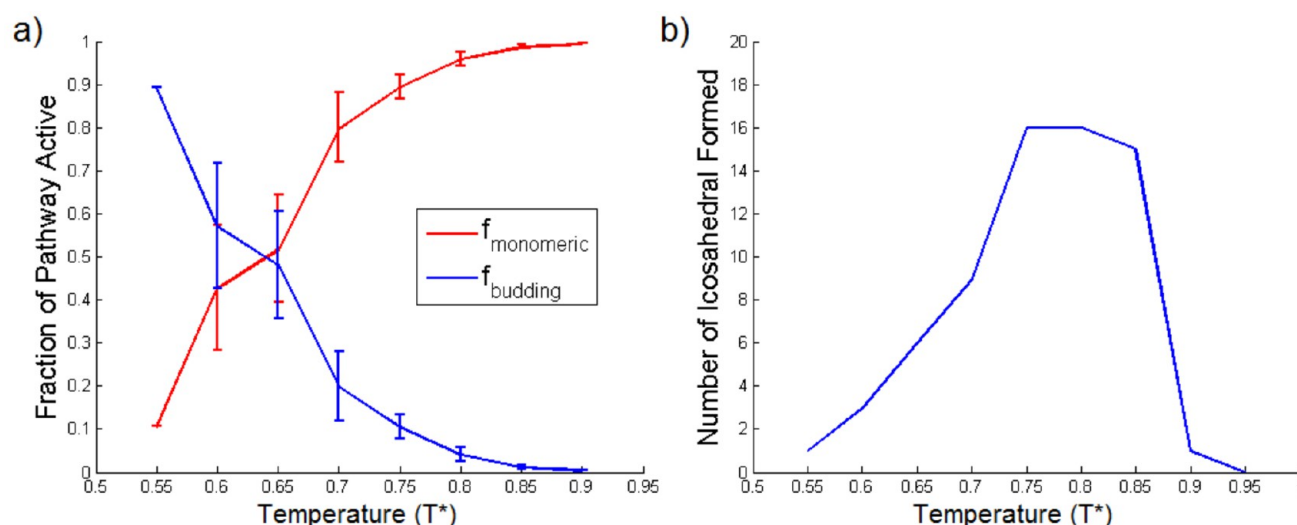


Figure 10. Icosahedral assembly fluxes. (a) Mean budding (red) and monomeric addition (blue) character of the icosahedral clusters observed to assemble at each temperature. Error bars represent the standard deviation in the icosahedral character averaged over all assembled icosahedra. (b) Total number of icosahedra assembled in the three independent simulations at each temperature.

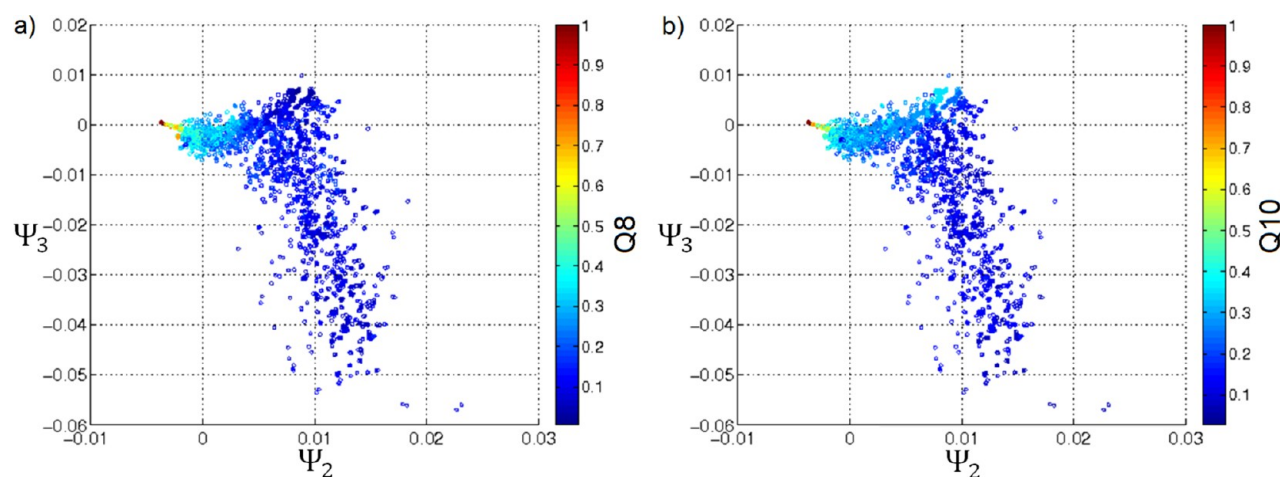


Figure 11. Two-dimensional embedding into $[\Psi_2, \Psi_3]$ of the 48 010 clusters harvested at all nine temperatures colored by (a) Q_8 and (b) Q_{10} .

temperatures disfavor the formation of large aggregates, Figure 6b), kinetics (high temperatures lead to faster rates, and low temperatures to kinetic trapping in a disordered liquid), and the topography of the free-energy surface (at high temperatures ($T^* \geq 0.85$), only the monomeric addition pathway is effectively open, whereas at lower temperatures both the monomeric addition and budding routes are available).

3.3. Physical Interpretation of Diffusion Map Coordinates. As previously observed here and elsewhere,³⁹ a limitation of the diffusion map approach is that order parameters describing the low-dimensional embedding (i.e., $\{\Psi_i\}$) come with no a priori physical interpretability. To provide post hoc understanding of the physical correspondence of these order parameters, we correlated Ψ_2 and Ψ_3 with candidate physical variables to identify two physical “bridge” variables showing strong correlation with each of these collective order parameters. In Figure 11, we color the combined temperature two-dimensional embeddings by the 8th and 10th order Steinhardt bond orientation order parameters, Q_8 and Q_{10} .⁷⁵

Q_8 approaches zero for clusters with icosahedral symmetry, enabling this variable to delineate between liquid or open

structures and the closed compact icosahedral clusters (Figure 11a). Ψ_2 exhibits strong correlation with Q_8 as measured by the Pearson correlation coefficient, $\rho(\Psi_2, Q_8) = 0.874$, with Q_8 falling from unity for dimers to zero for the terminal icosahedral aggregates. In particular, this parameter is capable of tracking the formation of icosahedral clusters along the monomeric assembly pathway but cannot distinguish between icosahedral and near-icosahedral aggregates with $Q_8 < 0.1$ and the disordered liquid phase.

Q_{10} is unity for dimers, approximately 0.363 for idealized icosahedra, and decreases toward zero as clusters become larger and more disordered. Accordingly, Q_{10} is a good variable by which to distinguish the liquid budding mechanism from monomeric assembly as illustrated in Figure 11b. Q_{10} therefore approximately tracks deviations in Ψ_3 , but the correlation is rather weak, with a Pearson correlation coefficient of $\rho(\Psi_3, Q_{10}) = 0.156$. Despite searching a variety of candidate variables (including principal moments of the gyration tensor, cluster size, mean number of bonds per particle in a cluster, Q_4 , Q_6 , Q_8 , Q_{10} , and Q_{12}), we were unable to find a physical variable possessing stronger correlation with Ψ_3 .

Our failure to find a physical variable strongly correlated with Ψ_3 underscores the difficulty in assigning simple physical interpretations to the diffusion map variables.^{39,62} While systematic procedures exist to search pools of candidate variables, the variable combinations so formed can often themselves be so complex as to obscure the underlying physical meaning.^{79,80} Furthermore, the diffusion map order parameters are themselves typically complex nonlinear collective variables, making the existence of a simple physical correspondence far from certain.^{39,61,62} While of unquestionable value in offering physical insight into the low-dimensional diffusion map embedding, the identity of diffusion map order parameters as good descriptors of the slow collective modes is independent of their physical interpretation,^{38,39,51,54} and the embeddings retain their validity and utility in the absence of this understanding. Furthermore, the Nyström extension^{39,52,83,86} enables direct embedding of new data points into the diffusion map embedding without possessing this physical correspondence, which facilitates efficient incorporation of new data and permits biased sampling directly in the diffusion map order parameters.⁸³

4. CONCLUSIONS

We have introduced a new approach to systematically infer self-assembly pathways and mechanisms from molecular simulations that is capable of identifying both *what* terminal aggregates are produced and *how* they are formed. Our approach is based on the nonlinear manifold learning approach known as the diffusion map,^{38,52,54,61} which we have previously used to identify the folding mechanisms of hydrocarbon chains and peptides.^{61,62} The key innovation of this work is the extension of this approach to treat many-body phenomena in which identical building blocks spontaneously assemble into structured terminal aggregates. In developing this generalization, we confronted a key technical challenge in defining structural similarities between clusters containing different numbers of identical particles in different configurations using graph-matching techniques.^{49,55} Our approach systematically identifies collective order parameters describing the long-time assembly dynamics of the system, furnishing a kinetically meaningful low-dimensional subspace in which to project simulation trajectories to reveal the underlying self-assembly pathways and mechanisms.

We have demonstrated our approach in applications to the self-assembly of anisotropic patchy particles with architectures designed to favor the assembly of tetrahedral and icosahedral clusters. Application of our approach to Brownian dynamics simulations of the self-assembly of icosahedral clusters reveals assembly to proceed by two distinct mechanisms: monomeric addition and budding from a disordered liquid. Both of these mechanisms were previously reported by Wilber et al. from viewing simulation trajectories and tracking of cluster size distribution,^{27,28} validating the capacity of our approach to systematically extract assembly mechanisms that were previously discernible only by visual inspection. Furthermore, by resolving the assembly pathways within a low-dimensional projection, our approach enabled us to explicitly compute the flux along each path and quantify their relative propensities as a function of temperature (interaction strength).

In an application to the particles designed to assemble tetrahedral clusters, our approach revealed the existence of two distinct assembly pathways to two distinct terminal aggregates: (i) chains of stacked interlocking dimers and tetramers and (ii)

chains of stacked interlocking trigonal planar trimers. Our approach revealed these chains to form by the condensation mechanism. While both pathways are active at low temperatures (strong particle interactions), we have shown that the trimer chain can be favored over the dimer/tetramer chain by elevating the temperature (decreasing the interaction strength). Despite designing our building block to favor the formation of tetrahedral clusters, we do not observe these structures to be uniquely favored at any temperature (interaction strength), suggesting that stabilization of tetrahedral clusters will require redesign on the anisotropic interactions. We observe that conducting and interpreting simulations of self-assembly present a powerful approach for unveiling the rich thermodynamic and kinetic features of the multibody assembly as a function of building block architecture and that naïve building block design considering only the desired aggregate is insufficient to guarantee its stability and/or accessibility.^{21,22,37}

These applications of our approach illustrate its capacity to systematically identify assembly pathways and mechanisms from molecular simulations, and provide fundamental insight into the assembly process. Furthermore, our approach furnishes a framework for quantifying how assembly conditions (e.g., temperature, pressure) or building block design (e.g., interaction strength and anisotropy, particle size and architecture) impact both the stable terminal aggregates and the mechanisms by which they assemble. Our exploration of the effect of temperature represents an exploration of the impact of one design variable (interaction strength), and in future work we plan to also explore the placement of the annular rings, as well as the role of hydrodynamic interactions, system size, and polydispersity in the building block architecture.

We observe that the kinetically meaningful collective variables inferred by the diffusion map present the “right” (i.e., kinetically meaningful) order parameters in which to construct free-energy landscapes integrating both the thermodynamic and kinetic aspects of assembly. In future studies, we shall recover these landscapes by performing umbrella sampling directly in the diffusion map order parameters.^{83,84} By quantifying the thermodynamic stability and kinetic accessibility of desired aggregates as a function of building block design within a single unified framework, we anticipate that these landscapes will offer a powerful vehicle for the rational design of building block architecture for assembling thermodynamically stable and kinetically accessible aggregates with desirable structural and/or functional properties. Ultimately, we envision that this computational design strategy may be profitably used to guide and inform experimental design, reducing reliance on laborious and expensive experimental trial and error and accelerating the search for building blocks capable of self-assembling valuable new materials such as antimicrobial nanostructures,⁸⁷ drug encapsulants,⁸⁸ and 3D tissue scaffolds.⁸⁹

■ ASSOCIATED CONTENT

● Supporting Information

Supplementary methods describing the modification of the IsoRank algorithm, determination of the soft-thresholding bandwidth in the diffusion maps, and determination of the diffusion map embedding dimensionality; Figures S1 and S2, illustrating the determination of the soft-thresholding bandwidth in the diffusion maps for the tetrahedral and icosahedral patchy particle systems, respectively; Figures S3 and S4, showing the diffusion map eigenvalue spectra for the

tetrahedral and icosahedral patchy particle systems, respectively. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon work supported by a National Science Foundation CAREER Award to A.L.F. (Grant No. DMR-1350008). The authors thank Ali Hajimirza and Abhijit Pujare for fruitful discussions on implementing the IsoRank algorithm.

ABBREVIATIONS

PCA - principal components analysis; WCA - Weeks–Chandler–Andersen; EMD - earth mover's distance; SVM - support vector machine

REFERENCES

- (1) Glotzer, S. C. *Science* **2004**, *306*, 419–420.
- (2) Cui, H.; Webber, M. J.; Stupp, S. I. *Pept. Sci.* **2010**, *94*, 1–18.
- (3) Matson, J. B.; Zha, R. H.; Stupp, S. I. *Curr. Opin. Solid State Mater. Sci.* **2011**, *15*, 225–235.
- (4) Glotzer, S. C.; Solomon, M. J. *Nat. Mater.* **2007**, *6*, 557–562.
- (5) Whitesides, G. M.; Grzybowski, B. *Science* **2002**, *295*, 2418–2421.
- (6) Koga, T.; Higuchi, M.; Kinoshita, T.; Higashi, N. *Chem.—Eur. J.* **2006**, *12*, 1360–1367.
- (7) Glotzer, S. C.; Solomon, M. J.; Kotov, N. A. *AIChE J.* **2004**, *50*, 2978–2985.
- (8) Whitesides, G. M.; Boncheva, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4769–4774.
- (9) Solomon, M. J. *Curr. Opin. Colloid Interface Sci.* **2011**, *16*, 158–167.
- (10) Chen, Q.; Bae, S. C.; Granick, S. *Nature* **2011**, *469*, 381–384.
- (11) Ning, H.; Mihi, A.; Geddes, J. B.; Miyake, M.; Braun, P. V. *Adv. Mater.* **2012**, *24*, OP153–OP158.
- (12) Zhao, X.; Pan, F.; Xu, H.; Yaseen, M.; Shan, H.; Hauser, C. A. E.; Zhang, S.; Lu, J. R. *Chem. Soc. Rev.* **2010**, *39*, 3480–3498.
- (13) Meng, Q.; Kou, Y.; Ma, X.; Liang, Y.; Guo, L.; Ni, C.; Liu, K. *Langmuir* **2012**, *28*, 5017–5022.
- (14) Zhang, S.; Marini, D. M.; Hwang, W.; Santoso, S. *Curr. Opin. Chem. Biol.* **2002**, *6*, 865–871.
- (15) Yoo, J.-W.; Irvine, D. J.; Discher, D. E.; Mitragotri, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 521–535.
- (16) Zlotnick, A.; Ceres, P.; Singh, S.; Johnson, J. M. *J. Virol.* **2002**, *76*, 4848–4854.
- (17) Sticht, J.; Humbert, M.; Findlow, S.; Bodem, J.; Muller, B.; Dietrich, U.; Werner, J.; Krausslich, H.-G. *Nat. Struct. Mol. Biol.* **2005**, *12*, 671–677.
- (18) Nie, Q.; Du, X.-G.; Geng, M.-Y. *Acta Pharmacol. Sin.* **2011**, *32*, 545–551.
- (19) Gazit, E. *FEBS J.* **2005**, *272*, 5971–5978.
- (20) Scherzinger, E.; Sittler, A.; Schweiger, K.; Heiser, V.; Lurz, R.; Hasenbank, R.; Bates, G. P.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4604–4609.
- (21) Jankowski, E.; Glotzer, S. C. *J. Phys. Chem. B* **2011**, *115*, 14321–14326.
- (22) Jankowski, E.; Glotzer, S. C. *Soft Matter* **2012**, *8*, 2852–2859.
- (23) Pawar, A. B.; Kretschmar, I. *Macromol. Rapid Commun.* **2010**, *31*, 150–168.
- (24) Wang, Y.; Wang, Y.; Breed, D. R.; Manoharan, V. N.; Feng, L.; Hollingsworth, A. D.; Weck, M.; Pine, D. J. *Nature* **2011**, *491*, 51–55.
- (25) Arpin, K. A.; Mihi, A.; Johnson, H. T.; Baca, A. J.; Rogers, J. A.; Lewis, J. A.; Braun, P. V. *Adv. Mater.* **2010**, *22*, 1084–1101.
- (26) Zhang, Z.; Keys, A. S.; Chen, T.; Glotzer, S. C. *Langmuir* **2005**, *21*, 11547–11551.
- (27) Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. J. *Chem. Phys.* **2007**, *127*, 085106.
- (28) Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Lewis, A. C. F. *J. Chem. Phys.* **2009**, *131*, 175102.
- (29) Hagan, M. F.; Chandler, D. *Biophys. J.* **2006**, *91*, 42–54.
- (30) Aniaeyei, S. E.; DuFort, C.; Kao, C. C.; Dragnea, B. J. *Mater. Chem.* **2008**, *18*, 3763–3774.
- (31) Steinmetz, N.; Lin, T.; Lomonosoff, G.; Johnson, J. Structure-Based Engineering of an Icosahedral Virus for Nanomedicine and Nanotechnology. In *Viruses and Nanotechnology*; Manchester, M., Steinmetz, N., Eds.; Springer: Berlin Heidelberg, 2009; Vol. 327, pp 23–58.
- (32) Boncheva, M.; Whitesides, G. M. *Dekker Encycl. Nanosci. Nanotechnol.* **2004**, 287–294.
- (33) Hagan, M. F. *J. Chem. Phys.* **2009**, *130*, 114902.
- (34) Hagan, M. F. *Phys. Rev. E* **2008**, *77*, 051904.
- (35) Sciortino, F.; Giacometti, A.; Pastore, G. *Phys. Rev. Lett.* **2009**, *103*, 237801.
- (36) Bianchi, E.; Blaak, R.; Likos, C. N. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6397–6410.
- (37) Hormoz, S.; Brenner, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5193–5198.
- (38) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7426–7431.
- (39) Ferguson, A. L.; Panagiotopoulos, A. Z.; Kevrekidis, I. G.; Debenedetti, P. G. *Chem. Phys. Lett.* **2011**, *509*, 1–11.
- (40) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002.
- (41) Zhang, Z.; Glotzer, S. C. *Nano Lett.* **2004**, *4*, 1407–1413.
- (42) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (43) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237–5247.
- (44) Berthelot, D. C. R. *Hebd. Seances Acad. Sci.* **1898**, *126*, 1703–1706.
- (45) Glotzer Group. *Shifted Lennard-Jones Pair Force*, http://codeblue.umich.edu/hoomd-blue/doc/classroomd__script_1_1pair_1_1slj.html, (accessed January 11, 2014).
- (46) Anderson, J. A.; Lorenz, C. D.; Travesset, A. J. *Comput. Phys.* **2008**, *227*, 5342–5359.
- (47) Nguyen, T. D.; Phillips, C. L.; Anderson, J. A.; Glotzer, S. C. *Comput. Phys. Commun.* **2011**, *182*, 2307–2313.
- (48) Miller, T. F.; Eleftheriou, M.; Pattnaik, P.; Ndirango, A.; News, D.; Martyna, G. J. *J. Chem. Phys.* **2002**, *116*, 8649–8659.
- (49) Lauck, F.; Helms, V.; Geyer, T. J. *Chem. Theory Comput.* **2009**, *5*, 641–648.
- (50) Tarjan, R. *SIAM J. Comput.* **1972**, *1*, 146–160.
- (51) Nadler, B.; Lafon, S.; Coifman, R. R.; Kevrekidis, I. G. Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker-Planck Operators. In *Advances in Neural Information Processing Systems 18: Proceedings of the 2005 Conference (Neural Information Processing)*; The MIT Press, Cambridge, MA, 2006; pp 955–962.
- (52) Sunday, B. E.; Haataja, M.; Kevrekidis, I. G. *Phys. Rev. E* **2009**, *80*, 031102.
- (53) Keys, A. S.; Iacovella, C. R.; Glotzer, S. C. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 263–285.
- (54) Coifman, R. R.; Lafon, S. *Appl. Comput. Harmonic Anal.* **2006**, *21*, 5–30.
- (55) Singh, R.; Xu, J.; Berger, B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12763–12768.
- (56) Beltran-Villegas, D. J.; Sehgal, R. M.; Maroudas, D.; Ford, D. M.; Bevan, M. A. *J. Chem. Phys.* **2012**, *137*, 134901.

- (57) Keys, A. S.; Iacovella, C. R.; Glotzer, S. C. *J. Comput. Phys.* **2011**, *230*, 6438–6463.
- (58) Rubner, Y.; Tomasi, C.; Guibas, L. J. A metric for distributions with applications to image databases. *Sixth International Conference on Computer Vision*, 1998. **1998**, 59–66.
- (59) Littlejohn, R. G.; Reinsch, M. *Rev. Mod. Phys.* **1997**, *69*, 213–276.
- (60) Conte, D.; Foggia, P.; Sansone, C.; Vento, M. *Int. J. Pattern Recognit. Artif. Intell.* **2004**, *18*, 265–298.
- (61) Ferguson, A. L.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Kevrekidis, I. G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13597–13602.
- (62) Ferguson, A. L.; Zhang, S.; Dikiy, I.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Link, A. J. *Biophys. J.* **2010**, *99*, 3056–3065.
- (63) Das, P.; Moll, M.; Stamati, H.; Kavraki, L. E.; Clementi, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9885–9890.
- (64) Plaku, E.; Stamati, H.; Clementi, C.; Kavraki, L. E. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 897–907.
- (65) Kentsis, A.; Gindin, T.; Mezei, M.; Osman, R. *PLoS ONE* **2007**, *2*, e446.
- (66) Zheng, W.; Qi, B.; Rohrdanz, M. A.; Caflisch, A.; Dinner, A. R.; Clementi, C. *J. Phys. Chem. B* **2011**, *115*, 13065–13074.
- (67) Zheng, W.; Vargiu, A. V.; Rohrdanz, M. A.; Carloni, P.; Clementi, C. *J. Chem. Phys.* **2013**, *139*, 145102.
- (68) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. *J. Chem. Phys.* **2011**, *134*, 124116.
- (69) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.
- (70) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press, New York, 2001.
- (71) Coifman, R.; Kevrekidis, I.; Lafon, S.; Maggioni, M.; Nadler, B. *Multiscale Model. Simul.* **2008**, *7*, 842–864.
- (72) Tenenbaum, J. B.; de Silva, V.; Langford, J. C. *Science* **2000**, *290*, 2319–2323.
- (73) Roweis, S. T.; Saul, L. K. *Science* **2000**, *290*, 2323–2326.
- (74) Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. Neural Inform. Process. Syst.* **2002**, *14*, 585–591.
- (75) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. *Phys. Rev. B* **1983**, *28*, 784–805.
- (76) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 6896–6906.
- (77) ten Wolde, P.; Ruiz-Montero, M.; Frenkel, D. *J. Chem. Phys.* **1996**, *104*, 9932–9947.
- (78) Filion, L.; Hermes, M.; Ni, R.; Dijkstra, M. *J. Chem. Phys.* **2010**, *133*, 244115.
- (79) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125*, 054108.
- (80) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769–6779.
- (81) Coifman, R.; Shkolnisky, Y.; Sigworth, F.; Singer, A. *IEEE Trans. Image Process.* **2008**, *17*, 1891–1899.
- (82) Salvador, S.; Chan, P. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. *16th IEEE International Conference on Tools with Artificial Intelligence*, 2004. **2004**, 576–584.
- (83) Laing, C. R.; Frewen, T. A.; Kevrekidis, I. G. *Nonlinearity* **2007**, *20*, 2127.
- (84) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (85) Choudhury, N.; Pettitt, B. M. *J. Phys. Chem. B* **2006**, *110*, 8459–8463.
- (86) Baker, C. T.; Baker, C. *The Numerical Treatment of Integral Equations*; Clarendon Press: Oxford, U.K., 1977; Vol. 13.
- (87) Engler, A. C.; Wiradharma, N.; Ong, Z. Y.; Coady, D. J.; Hedrick, J. L.; Yang, Y.-Y. *Nano Today* **2012**, *7*, 201–222.
- (88) Christian, D. A.; Tian, A.; Ellenbroek, W. G.; Levental, I.; Rajagopal, K.; Janmey, P. A.; Liu, A. J.; Baumgart, T.; Discher, D. E. *Nat. Mater.* **2009**, *8*, 843–849.
- (89) Zhang, S.; Gelain, F.; Zhao, X. *Semin. Cancer Biol.* **2005**, *15*, 413–420.