Dynamic remodelling of disordered protein aggregates is an alternative pathway to achieve robust self-assembly of nanostructures

Alia P. Schoen,ad Nicholas Cordella,bd Shafigh Mehraeen,c Manickam Adhimoolam Arunagirinathan,ad Andrew J. Spakowitzbd and Sarah C. Heilshorn*a

c © The Royal Society of Chemistry 2013

Clathrin is a naturally evolved protein that robustly assembles and disassembles into nanoscale spherical cages. This ability to reorganize in a highly dynamic fashion makes clathrin an attractive model system to study the kinetic and thermodynamic principles of biomolecular self-assembly. Through a combination of experimental and computational approaches, we demonstrate that competition between weak non-specific and specific reversible interactions can dictate the initial pathway of the assembly process, yet the final assembled structures are not sensitive to this competition. We conclude that the relative strengths of non-specific and specific interactions control clathrin assembly at short time scales resulting in either disordered protein aggregates or regularly structured assemblies. However with sufficient time for remodeling, the final assembled structure is robustly formed due to geometric constraints arising from specific molecular recognition events. These data provide insight into naturally evolved biological assembly processes and guidance for the design of engineered systems to achieve robust assembly.

1 Introduction

The self-assembly of biomolecules is ubiquitous in nature and leads to the formation of intricate structures providing a wide variety of structural, chemical, and physiological functionality. These structures form from complex nanoscale building blocks and with remarkable reproducibility making biological self-assembly an attractive processing route in a variety of applications. DNA,1,2 virus-like particles,3–5 bacterial membranes,6,7 and proteins8–10 show promise as biotemplates to synthesize and organize inorganic materials at the nanoscale. An understanding of the principles that govern complex biological self-assembly pathways will lead to predictive control and enhanced utility of self-assembled biomolecules in nanoscale applications.

Clathrin is a protein with three legs of about 50 nm in length that meet at a central vertex, forming a triskelion structure.11 Clathrin self-associates to form a variety of structures. While the exact molecular-level details are as yet unknown, this self-assembly is stabilized by multiple weak leg-leg interactions.12 In its natural environment inside a cell, clathrin plays a key role in the dynamic reshaping of the cell membrane during endocytosis.11 The individual triskelia localize at the lipid membrane and dynamically organize to create curved pits that pinch off to form spherical cages around lipid vesicles. The clathrin then disassembles and recycles back to the membrane to begin the process again. The highly dynamic nature of the assembly process makes clathrin an attractive model system to study the kinetic and thermodynamic principles of biological self-assembly.

Central to the ability to robustly form ordered self-assembled structures is the capability to overcome local energetic minima that arise during the assembly process. Although strong specific interactions can be used to achieve self-assembly of specific tailored structures, this design strategy can lead to kinetically trapped irregular structures that are incapable of reorganizing on experimentally realizable timescales. Alternatively, the prevalence of specific interactions that are weak and reversible allows dynamic reorganization, which may enable the system to explore a wider variety of geometric configurations thereby improving the probability that the thermodynamically desired structure will be formed. In addition, self-assembling molecules in nature, such as clathrin, may combine local and specific order-determining interactions with delocalized, non-specific attraction and repulsion to achieve robust assembly.

To explore the competing roles of specific and non-specific interactions during biological self-assembly, we combine
experimental and computational approaches to systematically study the formation of the most commonly observed in vitro clathrin assembly, spherical cages. Dynamic light scattering (DLS) is used in combination with cryo transmission electron microscopy (TEM) to assess time dependent clathrin self-assembly for various environmental pH, which tunes the surface charge and hydrophobicity of the protein. We have previously developed a detailed, mesoscale theoretical model of clathrin at a 2D surface. Here we extend this model to 3D and use it in Brownian dynamics simulations to study the effects of competing non-specific and specific reversible interactions on clathrin self-assembly and remodeling in solution. Our studies demonstrate that at least two distinct kinetic routes of self-assembly exist for clathrin, both resulting in similar spherical cage architectures. The timescale of assembly is dictated by the relative strengths of non-specific and specific interactions, which are tuned by pH manipulation. These results provide insight into biological mechanisms of dynamic self-assembly and provide guidance in the future design of engineered self-assembly systems.

2 Results
Clathrin protein is isolated from bovine brain tissue using differential centrifugation and size exclusion chromatography as previously reported. Purified clathrin triskelia are induced to self-assemble in 100 mM 2-morpholinoethanesulfonic acid (MES) with pH values between 5.1 and 6.9 and monitored in situ by DLS. The DLS size distribution data indicate three main states of clathrin assembly (Fig. 1). Based on TEM analysis of the protein solutions, we attribute these three states to different discrete structures: monomeric clathrin, assembled cages, and disordered aggregates (Fig. 1a–c, respectively). In addition, certain solutions exhibit multiple peaks within the light scattering traces, which is consistent with a mixture of species that might occur during a dynamic self-assembly process (Fig. 1d).

To investigate the kinetics of self-assembly, clathrin assembly reactions are monitored by DLS over 20 minutes with measurements every minute. Taking measurements after 1 day and 5 days assesses long-term stability of the assembled structures. In the first 20 minutes of assembly the process is highly dynamic; size distribution peaks are broad and multiple peaks are seen often. The peaks after 1 day are significantly sharper, indicating a narrower size distribution of structures in solution (Fig. 2a). We examine pH-dependent trends by plotting the size of the dominant peak to identify the major species and its evolution with time (Fig. 2b). The data indicate two distinct kinetic routes of self-assembly. At pH values below the isoelectric point (IEP) of clathrin, pH 5.8, the protein quickly aggregates in solution and subsequently adopts a cage-like signature in the DLS data. In contrast, at pH values above the IEP, the assembly proceeds monotonically from monomer to cage. At a pH near the IEP, where the net charge of the protein is near zero, cage structures form monotonically but are unstable and aggregate over time. These data suggest that surface charge plays an important role in regulating dynamic self-assembly and reorganization of discrete structures and disordered aggregates. However while the IEP strictly describes net surface

Fig. 1 Representative hydrodynamic-radius distributions from DLS (left), representative TEM images (middle), and representative simulations (right) of clathrin protein solutions containing (a) monomers, (b) assembled cages, (c) disordered aggregates, and (d) a mixture of assembled cages and disordered aggregates.
charge, a reduction in pH also increases clathrin hydrophobicity,\textsuperscript{16} so we are not able to attribute these effects exclusively to changes in surface charge.

Next, cryo TEM is used to examine the structures that form along the two distinct kinetic pathways above (pH 6.0) and below (pH 5.1) the IEP. After 20 minutes of assembly, spherical cages are observed in the sample at pH 6.0, as expected from the presence of a small broad peak near 100 nm in the DLS data (Fig. 3a). These structures are stable and still observable after 4 days (Fig. 3c). Individual cages appear to have subtle differences in size, which is consistent with the observed heterogeneity of clathrin structures that arise within a cell during endocytosis.\textsuperscript{11} Similar heterogeneity has been documented experimentally in other biological self-assembly systems such as viral capsid structures.\textsuperscript{17} In contrast, large disordered aggregates are observed in the sample held at pH 5.1 for 20 minutes, corresponding to the presence of a peak between 400 and 500 nm in the DLS data (Fig. 3b). After 4 days, these same pH 5.1 samples exhibit formation of ordered spherical cage structures, consistent with the observed DLS peak shift at longer times (Fig. 3d). These data demonstrate that large scale remodeling of the clathrin aggregates occurs over time, resulting in final structures that are similar in size and shape to the cages that grow monotonically at higher pH (Fig. 3e). Therefore, both kinetic routes of assembly ultimately result in similar final assembled states, suggesting that the mechanism of clathrin assembly has evolved to achieve robustness. This apparent robustness of assembly into cage-like structures may be an essential feature for clathrin to perform critical functions in a wide range of organisms.\textsuperscript{18}

Brownian dynamics simulations provide insight into the mechanisms that drive assembly through the various experimentally observed pathways. Our simplified model for this self-assembly process treats the individual triskelia as three-legged pinwheels, whose deformation energies are described by harmonic terms that resist stretching and twisting of bonds, as well as the distortion of the angles at which the legs emanate from the central hub. Any unbound legs are assumed to adopt a

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig2}
\caption{(a) Kymographs of DLS data showing change in hydrodynamic radius distribution over time for a range of pH values. Hydrodynamic radius is shown on the x-axis, time increases in increments of 1 minute from top to bottom, and color indicates the volumetric concentration. Data for 1 day and 5 days time points are shown in separate bars. (b) Plot of hydrodynamic radius over time for the most populous species in solutions of various pH.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig3}
\caption{(a–d) Cryo TEM images of clathrin assemblies at pH 6 (a and c) and pH 5.1 (b and d) after 20 minutes (a and b) and 4 days (c and d) of assembly. Representative spherical cages and aggregates are outlined in blue and red, respectively. (e) Dot plot showing the major axis distribution of ellipses fitted to clathrin assemblies at pH 6.0 (circles) and pH 5.1 (triangles) after 20 minutes (open) and 4 days (filled). Solid line represents the mean value (n = 63, 169, 64, and 54).}
\end{figure}
minimum energy configuration, so elastic strain on our clathrin assemblies only results from translocation and rotation of hubs that are bound to others. In this way, the state of the system is fully defined by the location and orientation of the triskelion hubs. This formulation enables us to simulate timescales that allow for many clathrin binding and aggregation events in a system of several hundred triskelia, which would not be possible had our model included atomistic detail of this large, 650 kDa protein complex.

Our model triskelia interact with one another through both non-specific and specific interactions. An unbound triskelion leg can form a specific bond with an unbound leg of another triskelion, resulting in an energetic benefit of $-\epsilon$. The $K_d$ for this interaction has been reported to be $\sim 100 \mu M$ (i.e. $\sim 10 \ k_BT$). The affinity $\epsilon$ is set to $15 \ k_BT$ for all of our simulations which results in closed cage formation on the timescale of the simulations. The triskelia interact non-specifically through a Lennard-Jones potential, given by

$$E_{LJ,ij} = \frac{\sigma}{r_{ij}^{12}} - \frac{2\sigma}{r_{ij}^6}$$

where $r_{ij}$ is the separation distance between the $i$th hub and the $j$th hub, $\sigma$ is the value of $r_{ij}$ that minimizes $E_{LJ}$ (i.e., the distance at which non-specific interactions are most favorable), and $v_0$ is the potential well depth (i.e., the strength of non-specific interactions). By changing the value of $v_0$ in our model, we are able to observe the effects of shifting the balance between specific and non-specific driving forces on clathrin self-assembly pathways.

Altering the strength of non-specific interactions through $v_0$ results in starkly different assembly pathways. Over short simulated timescales, systems with $v_0 = 1 \ k_BT$ showed a tendency to aggregate into large disordered structures, while systems with $v_0 = 0.1 \ k_BT$ assembled into well-defined cages. We scale our simulation time by the Brownian time step $T$, which is the time required for a triskelion to one clathrin leg length on average. As shown by our simulation snapshots, at $v_0 = 0.1 \ k_BT$, a random spatial distribution of monomers (with free legs identified in green) at $t = 0$ bind to one another over time to form cage-like structures that are almost completely held together through specific leg–leg interactions (Fig. 4a). This is in contrast to the dynamic aggregation of monomers into larger aggregated clusters in simulations that use $v_0 = 1 \ k_BT$ (Fig. 4b).

We quantify the sizes of our simulated clathrin assemblies by calculating their hydrodynamic radii, $R_H$ (Fig. 4c and d). These values are nearly identical to the assemblies’ radii of gyration. $R_H$ is obtained using two distinct definitions of “clathrin clusters”. When the cluster is defined as a structure connected through specific leg–leg interactions only, the hydrodynamic radii for both values of $v_0$ monotonically approach $\sim 50$ nm over time. However, by defining a cluster based on the proximity of triskelion hubs (i.e., individual hubs are separated by less than 1.25 leg lengths regardless of specific bond formation), the $R_H$ behavior over time is different between the two values of $v_0$. In the simulation with weaker non-specific interactions ($v_0 = 0.1 \ k_BT$), definition of cluster size has little impact on calculated $R_H$. Both definitions lead to cluster sizes that appear to plateau over time to 50–60 nm, consistent with
the formation of discrete spherical cages (Fig. 4a and c). In contrast, increasing the strength of non-specific interactions \( \nu_0 = 1 K_B T \) predicts a steady growth of \( R_M \) for clusters defined by hub proximity (Fig. 4d). Already at the end of this simulation time, these clusters are 100% larger than their specifically bound counterparts, and the value has not yet reached a plateau.

The observed robust assembly behavior for specifically bound clusters is derived from the geometry of a clathrin triskelion. Leg length and pucker angle of the triskelion dictate the resulting cage size rather than interaction strength. We also observe this experimentally in both DLS and TEM which show that cage sizes are similar after assembly in different pH conditions. These results suggest that geometrical requirements for formation of specific bonds between legs play a significant role in structure determination.

It is also notable that the aggregates formed when \( \nu_0 = 1 K_B T \) consist of clumps of partial cage-like structures that individually resemble the clusters in the early stages of simulations with \( \nu_0 = 0.1 K_B T \). This behavior shows that the specific leg-leg interactions exert significant control over short-time clathrin kinetic assembly, even when nonspecific interactions dictate the large-scale assembly patterns.

3 Discussion

Often during self-assembly processes, deviation from a single deterministic pathway leads to disordered aggregation and kinetic traps that are difficult to exit. Classical nucleation and growth theory predicts a monotonic increase of cluster size once the critical nucleus size is overcome. Subsequent cluster growth can lead to secondary mechanisms of reorganization including coalescence and Ostwald ripening; however, these secondary growth mechanisms would generally lead to the volume average cluster size increasing monotonically with time. The dynamics of cluster formation can exhibit non-monotonic behavior particularly in finite-sized systems that exhibit kinetic intermediates that exhaust the solution of the fundamental subunits, and subsequent reorganization of these clusters can exhibit a decrease in cluster size. Here, our solution phase assembly is not limited by the clathrin number yet exhibits non-monotonic behavior. DLS and cryo-TEM analysis of clathrin self-assembly suggests that in addition to following a monotonic growth pathway, clathrin is also able to achieve ordered nanostructures through a distinct pathway by remodeling large disordered aggregates.

When aggregation occurs, the local concentration of clathrin increases dramatically. Therefore, the availability of specific binding sites is increased and the clathrin legs can readily find new binding partners if rearrangement of bonds is allowed. It is important to note that individual clathrin leg-leg interactions have been reported to be quite weak; the \( K_d \) has been reported to be on the order of 100 µM \( (i.e. \sim 10 K_B T) \). The free energy gain from this weak binding is likely not sufficient to drive rearrangement of a single triskelion in a disordered aggregate; however, weak binding would be necessary to allow bond motion in order to leave the kinetically frustrated aggregate state on experimentally realizable timescales.

Previous work on modeling clathrin assembly in two dimensions shows that defects in an otherwise regular flat honeycomb lattice interact with each other over a very long distance which can cause them to attract one another and subsequently annihilate. For closed three-dimensional structures, the same long range elastic interaction between defects may not cause long-range attraction to occur; however, weak reversible interactions between clathrin legs permit defects to diffuse due to thermal reorganization. Because both non-specific and specific binding are weak, any interactions that do not satisfy the geometrical requirements for formation of a thermodynamically favorable spherical cage can be rearranged until a minimum elastic strain state is reached.

Our computational modeling of three-dimensional assembly in solution implicates the strength of non-specific interactions between clathrin monomers as a determining factor of clathrin-structure size and order on short timescales. The large variations in the structure sizes that we observe experimentally at different solution conditions align with these predictions. Decreasing pH corresponds to an increase in non-specific interaction strength due to a number of simultaneous changes in the protein including changes in surface charge and hydrophobicity. At pH above the IEP, clathrin carries a net negative charge resulting in a non-specific repulsion between monomers. However, if the short lengthscale specific interactions between legs can overcome this repulsion, this would lead to the observed cage formation behavior at higher pH. At pH below the IEP, it has been shown that clathrin becomes increasingly hydrophobic. This corresponds to the exact regime where our data indicate non-specific aggregation occurs. This increase in hydrophobicity likely creates an attractive force strong enough to counteract any electrostatic repulsion from the net positive surface charge, resulting in fast aggregation of clathrin monomers before specific associations can be formed. Our 3D simulations only represent very short timescales due to computational limitations, so we do not directly observe the modeled clathrin aggregates annealing into spherical cage-like structures. However, such behavior in our model system is fully expected to occur as the dynamic simulations slowly readjust the positions and connectedness of the clathrin structures to reduce the free energy of the system.

In typical engineered self-assembling systems, non-specific attraction is usually avoided to prevent flocculation. We suggest that weak delocalized attraction may be used as a way to bring monomers together, provided a sufficient driving force for rearrangement into the proper geometry exists. This phenomenon has been demonstrated before in a two-dimensional colloidal system where tuning of hydrophobic attraction along with geometrical constraints on binding sites led to the formation of a kagome lattice structure from disordered intermediates. In a comparable two-dimensional protein system, collagen assembly on mica surfaces has been shown to be sensitive to the relative strength of non-specific collagen-mica interaction and specific collagen-collagen interaction, where geometrically ordered collagen bundling due to specific
interactions follows a dense disordered state mediated by non-specific attraction to the mica surface. Additionally, two-dimensional S-layer protein assembly on supported lipid bilayers has been shown to proceed via a dense amorphous state before crystallization, and the balance of specific and non-specific interactions has been shown to profoundly affect the assembly pathway. We show that these principles also hold true when assembling protein in solution. Tuning the relative strengths of the non-specific and specific bonding can result in distinct kinetic pathways to the same nanoscale assembled structure. This implies that when designing biomimetic self-assembling systems, non-specific interactions can be used to promote a local increase in available binding sites while weak specific interactions are used to determine final order of the structure.

4 Conclusions

We demonstrate that clathrin self-assembles into spherical cage structures via two kinetically distinct routes. One pathway involves monotonic growth of the cages mediated primarily by specific leg-leg interactions. The second pathway begins with disordered aggregation due to non-specific interactions followed by dynamic remodeling into discrete ordered cages as the more favorable specific leg-leg interactions are formed. These results suggest that in designing de novo self-assembling nanostructures, non-specific interactions, such as hydrophobic condensation, can be leveraged as additional parameters to achieve robust assembly, and use of multiple weak specific interactions will allow escape from disordered kinetically trapped species while still encoding a desired ordered structure.

5 Methods

Clathrin purification

Bovine brain tissue is blended in an equal volume of HKM buffer (25 mM HEPES, 125 mM potassium acetate, 5 mM magnesium acetate, 1 mM DTT, pH 7.4), homogenized, and spun at 5750 × g for 20 minutes. The supernatant is collected and spun at 43 000 × g for 40 minutes. The pellets are resuspended in HKM, mixed with an equal volume of 12.5% w/v Ficoll and 12.5% w/v sucrose in HKM, sonicated for 1 minute, and spun at 25 000 × g for 20 minutes. The supernatant is collected, combined with 3 volumes of HKM, and spun at 35 000 × g for 60 minutes. Each resulting pellet of clathrin-coated vesicles is resuspended in HKM for storage. For clathrin monomer isolation, vesicles are resuspended in dissociation buffer (1 M Tris pH 7.0, 1 mM EDTA) then spun down at 22 000 × g for 20 minutes before loading the supernatant onto a Sepharose CL-4B gel filtration column. Fractions are analyzed by SDS-PAGE, pooled, and concentrated before storage in dissociation buffer.

Dynamic light scattering and clathrin assembly

Dynamic light scattering is carried out in a Malvern Zetasizer Nano in backscatter mode. Clathrin protein is prepared for assembly by dialysis into 2 mM Tris buffer at pH 7.0 and made into a stock solution of 50 µg mL⁻¹. 90 mL of stock solution is scanned prior to assembly as an initial reference curve. 10 mL of appropriate assembly buffer is added to the sample, mixed, and immediately inserted into the DLS for scanning. Scans are taken for 20 minutes at 1-minute intervals, overnight, and 5 days later. Samples are stored capped and at room temperature. For the pH series, the assembly buffers are 1 M MES at pH 5.1, 5.4, 5.7, 6.0, 6.3, 6.6, or 6.9.

Cryo-TEM sample preparation and imaging

Cryo-TEM grids are prepared using an FEI Vitrobot Mark IV controlled environment vitrification system (CEVS). A Quantifoil holey carbon R2/4 200 mesh grid is glow discharged using a Harrick Plasma PDC-32G for 30 s at 500 mtorr, low setting. 3 mL of sample is applied to the carbon side of the grid inside the CEVS chamber maintained at 22 °C and 100% relative humidity. After 5–10 s, excess sample is blotted for 3 s with blot force 9 using Whatman Grade 597 filter paper to form a thin film of sample. The grid is plunged into liquid ethane to form a vitrified film of sample on the grid. The sample grid is transferred under liquid nitrogen vapor onto a Gatan cryo holder, which is stabilized at a temperature less than −175 °C, and then transferred inside the TEM and maintained at less than −175 °C during imaging. The sample is imaged in an FEI Tecnai F20 with field emission gun filament at 200 kV in low dose mode using ~10 to 20 e⁻/Å² electron dose and −5 to −8 µm defocus to obtain sufficient contrast.

Computational methods

Our model of clathrin triskelia extends our previous two-dimensional work to include the three-dimensional position and orientation of individual clathrin triskelia. We use a simplified elastic representation, in which each triskelion is defined by a hub from which three legs emanate in a pinwheel configuration. Unbound legs of nearby triskelia are capable of binding to one another, contributing an energetic benefit of −εb and bound legs are capable of unbinding with a corresponding energetic penalty. Our simulations use a value of ε = −15 kBT. A Brownian time step, τB, is defined as the time for which independent clathrin hubs diffuse a distance equal to one leg length r0. We run our simulations for a total of 343 Brownian time steps. The state of our system is completely defined by the spatial locations of the hubs ̂r i, their normal vectors ̂n i, and the bond connectivity network, which is characterized by the link indicator L ij and the leg index indicator λi j. The link indicator L ij is equal to 1 when hubs i and j are bound to one another and equal to 0 otherwise. If L ij = 1, the leg index indicator λi j gives the index of the leg on hub i (between 1 and 3) that connects hub j. If L ij = 0, the leg index indicator λi j = 0. When a triskelion is bound to other neighboring triskelia, the bound triskelia are deformable through four elastic modes, all of which obey a Hookean spring-like potential. These elastic modes include stretching of bound legs, twisting of bound legs, in-plane bending of legs away from their uniform radial distribution around the hub, and out-of-plane bending of legs away from their intrinsic pucker angle. The clathrin hubs also...
interact non-specifically via a Lennard-Jones potential. Correspondingly, the total energy for a system of N pinwheels in our model is given by

\[
E = -\epsilon \sum_{i=1}^{N} \sum_{j=i+1}^{N} L_{ij} + v_0 \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma}{r_{ij}} \right)^{6} \right] + k_{r} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (r_{ij} - r_0)^2 L_{ij} + k_{\phi} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_{ij}^2 L_{ij}
\]

The distance between hubs \(i\) and \(j\) is given by \(r_{ij}\), and the equilibrium bond length between hubs is \(r_0\). We use \(r_0 = 20\, \text{nm}\), which roughly corresponds to previously measured lengths between vertices of reassembled clathrin cages;\textsuperscript{24} although, the edge length is also measured to be as small as 15\,\text{nm} by some researchers.\textsuperscript{25} The Lennard-Jones potential is characterized by the potential well depth \(v_0\) and the distance of minimum non-specific energy \(\sigma\). All simulations use \(v_0 = r_0\) and we vary the value of \(v_0\). Elastic moduli for the stretching \(k_{s}\), twisting \(k_{t}\), in-plane bending \(k_{ipb}\), and out-of-plane bending \(k_{opb}\) are equal to 67 \(k_B T/\pi r_0^2\), 58 \(k_B T\), 36 \(k_B T\), and 36 \(k_B T\), respectively. These values are chosen to approximately correspond with the triskelia rigidities that are determined experimentally using leg-angle and leg-length distributions from electron micrographs.\textsuperscript{26,27}

The twisting energy is a function of the angle \(\phi_{ij}\) representing the misalignment of the normal vector components of hubs \(i\) and \(j\) that are orthogonal to their connecting bond. The calculated in-plane bending metric \(\chi_{ij}^k\) describes the contribution to the in-plane bending energy from the legs on hub \(i\) that are connected to hubs \(j\) and \(k\). The value of \(\chi_{ij}^k\) depends on the binding state of the third leg of hub \(i\). If the third leg of hub \(i\) is bound to another hub with index \(l\) (i.e. \(\lambda_l^i\) is not zero), the in-plane bending energy is strictly given by the angle between hubs \(i, j, k\), resulting in

\[
\chi_{ij}^k(\lambda_l^i, \lambda_j^k, \lambda_k^l) = 0 = \left( \frac{\phi_{ij}^{l,j,k} - \frac{2\pi}{3}}{2} \right)^2
\]

If the third leg of hub \(i\) is not bound to any other hub (i.e. \(\lambda_l^i\) is zero for all \(l\)), then that leg is assumed to adopt a configuration that minimizes in-plane bending energy between it and the other two legs, such that

\[
\chi_{ij}^k(\lambda_l^i, \lambda_j^k, \lambda_k^l) = 0 = \left( \frac{\phi_{ij}^{l,j,k} - \frac{2\pi}{3}}{2} \right)^2
\]

The component of the angle between legs \(\lambda_l^i\) and \(\lambda_j^k\) on hub \(i\) that is orthogonal to hub \(i\)'s normal vector \(\vec{n}_i\) is given by \(\phi_{jk}^{l,j,k}\) and is measured in the counterclockwise direction from leg 1 to leg 2, leg 2 to leg 3, or leg 3 to leg 1. Out-of-plane elastic energy is dictated by the difference between the clathrin’s intrinsic pucker angle \(\alpha_0\) and \(\alpha_i\), the angle between \(\vec{n}_i\) and the tangent vector of hub \(i\)'s leg that connects it to hub \(j\).

A combined Brownian dynamics and dynamic Monte Carlo simulation is used to predict the motion and binding processes of triskelia in solution. We simulate the behavior of 500 triskelia in a cubic, periodic box of edge length \(L = 20a\), using a 2nd order Runge-Kutta routine. The data presented in this paper represents the average of 15 independent simulations for each different set of parameters. Neglecting long-range hydrodynamic coupling between clathrin, we employ the following overdamped Langevin equations of motion to predict the dynamic behavior of \(\vec{r}_i\) and \(\vec{n}_i\):

\[
\vec{r}_i = \vec{r}_i - \vec{F}_t \frac{\vec{F}_t}{|\vec{F}_t|^2} - \vec{F}_{v} - \vec{F}_k - \vec{F}_{\text{int}} - \vec{F}_{\text{ext}} - \vec{F}_{\text{brown}},
\]

\[
\vec{n}_i = \vec{n}_i - \vec{T}_t \frac{\vec{F}_t}{|\vec{F}_t|^2} - \vec{T}_{v} - \vec{T}_k - \vec{T}_{\text{int}} - \vec{T}_{\text{ext}} - \vec{T}_{\text{brown}},
\]

The translational and rotational drag coefficients, \(\xi_t\) and \(\xi_n\), balance the potential-driven force and torque as well as the Brownian (i.e. thermal) force and torque, \(\vec{F}_t\) and \(\vec{T}_t\). The magnitudes of the thermal force and torque are chosen to obey the fluctuation–dissipation theorem

\[
\left\langle \vec{F}_t(t) \vec{F}_t(t') \right\rangle = -2k_B T \delta(t - t') \delta_{ij} I
\]

\[
\left\langle \vec{T}_t(t) \vec{T}_t(t') \right\rangle = -2k_B T \delta(t - t') \delta_{ij} I
\]

The binding and unbinding dynamics are predicted using a set of configuration-dependent rate equations. If \(L_0 = 0\), the binding dynamics are governed by the rate constant \(q_{b}(\lambda_l^i, \lambda_j^k, \lambda_k^l)\) for binding leg \(\lambda_l^i\) of hub \(i\) to leg \(\lambda_j^k\) of hub \(j\). Alternatively, if \(L_0 = 1\), the unbinding dynamics are governed by \(q_{u}(\lambda_l^i, \lambda_j^k, \lambda_k^l)\). These rate constants are continuously updated as the simulation evolves, such that

\[
q_b(i, \lambda_l^i | j, \lambda_j^k) = \frac{\omega}{1 + \exp \left[ -\beta \Delta E_b(i, \lambda_l^i | j, \lambda_j^k) \right]}
\]

\[
q_u(i, \lambda_l^i | j, \lambda_j^k) = \frac{q_u(i, \lambda_l^i | j, \lambda_j^k)}{\omega - q_u(i, \lambda_l^i | j, \lambda_j^k)}
\]

where \(\Delta E_b(i, \lambda_l^i | j, \lambda_j^k)\) is the change in energy of the entire system due to forming the \(i-j\) bond, calculated using eqn. (2). The reaction frequency \(\omega\) is a fundamental reaction rate that is independent of energy. These rate constants, as well as the forces and torques dictating motion, are calculated after every time integration step \(\Delta t\), based on the updates to the positions and connections that occur during that step. There are 4000 integration steps within each Brownian time step.

Between times \(t\) and \(t + \Delta t\), we perform a dynamic Monte Carlo simulation, based on the Gillespie algorithm,\textsuperscript{28} to determine the bonds that are made and broken during this period. This procedure starts by calculating the overall reaction rate constant

\[
q_{\text{tot}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ q_b(i, \lambda_l^i | j, \lambda_j^k) + q_u(i, \lambda_l^i | j, \lambda_j^k) \right]
\]

We then randomly select a number \(\gamma_1\) between 0 and 1, and if \(\gamma_1 > \exp(-q_{\text{tot}} \Delta t)\), a reaction is determined to occur. The exact
reaction that occurs is randomly selected based on the probabilities of each binding reaction \( p_b(i, j; i', j') = q_b(i, j; i', j') / q_{tot} \) and each unbinding reaction \( p_u(i, j; i', j') = q_u(i, j; i', j') / q_{tot} \). The time of the reaction is then determined by

\[
\tau' = t - \frac{1}{q_{tot}} \log \{ \gamma_2 [\exp(-q_{tot}\Delta t) - 1] + 1 \} \tag{12}
\]

where \( \gamma_2 \) is a random number between 0 and 1. The rate constants \( q_u \) and \( q_b \) are updated to reflect the new connectivity, and then the dynamic Monte Carlo process is repeated over the shorter time interval \( t + \Delta t - \tau' \), instead of \( \Delta t \). This is repeated over shorter and shorter steps until a value of \( \gamma_1 \) is chosen such that no reactions occur.

Our simulations take several days to run and represent approximately 100 ms of real time. However, experimental observations are not possible on that timescale. Therefore, we use our simulations to make qualitative predictions regarding the kinetic pathways of clathrin assembly, while choosing certain input parameters to ensure that we can observe this process within a reasonable computational time frame. Specifically, we set the energy-independent reaction frequency to 100 times higher than the density of our experimental scattering results, we compute the volume average hydrodynamic radius of the \( N_c \) individual clusters, such that

\[
R_H = \left( \frac{\sum_{m=1}^{N_c} R_{H,m}^4}{\sum_{m=1}^{N_c} R_{H,m}^3} \right)^{\frac{3}{4}} \tag{16}
\]

**Acknowledgements**

We thank Kelly Huggins for assistance with protein purification, and Tom Lampo, Nick Melosh and Seb Doniach for helpful discussions. We thank Dr Brian Gibbons for assistance with cryo TEM imaging and Roger Kornberg, Stanford University, and Yifan Cheng, University of California, San Francisco, for use of cryo TEM facilities. A. P. S. acknowledges support from a Stanford Bio-X fellowship and an Achievement Rewards for College Scientists Foundation (ARCS) Scholarship. This work is supported by the Department of Energy Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract DE-AC02-76SF00515.

**Notes and references**