Biophysical Journal
Article

Biophysical Society

# Clathrin Assembly Regulated by Adaptor Proteins in Coarse-Grained Models

Matteo Giani,[1,2,3] Wouter K. den Otter,[1,2,3,*] and Wim J. Briels[2,3,4]
[1]Multi Scale Mechanics, Faculty of Engineering Technology, [2]Computational BioPhysics, Faculty of Science and Technology, and [3]MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands; and [4]Forschungszentrum Jülich, Jülich, Germany

ABSTRACT The assembly of clathrin triskelia into polyhedral cages during endocytosis is regulated by adaptor proteins (APs). We explore how APs achieve this by developing coarse-grained models for clathrin and AP2, employing a Monte Carlo click interaction, to simulate their collective aggregation behavior. The phase diagrams indicate that a crucial role is played by the mechanical properties of the disordered linker segment of AP. We also present a statistical-mechanical theory for the assembly behavior of clathrin, yielding good agreement with our simulations and experimental data from the literature. Adaptor proteins are found to regulate the formation of clathrin coats under certain conditions, but can also suppress the formation of cages.

## INTRODUCTION

In eukaryotic cells, clathrin-mediated endocytosis is a major pathway for the internalization of cargo molecules such as hormones, receptors, transferrin, membrane lipids, and the occasional virus (1–9). The cargo molecules are collected and sorted in a clathrin-coated pit, which subsequently evolves into an encapsulating clathrin-coated vesicle. These coats arise through a self-assembly or polymerization process of clathrin proteins against the cytoplasmic face of cellular membranes. The clathrin protein has a peculiar shape with three long curved legs (see Fig. 1), which allows it to bind with many partners into a wide range of polyhedral cages, as well as to bind accessory proteins that assist at various stages of the endocytosis process (10–15). Although clathrin is a major component and the namesake of clathrin-coated pits and clathrin-coated vesicles, it does not bind directly to either the membrane or the cargo. These are the tasks of so-called adaptor proteins, which often are active only at specific membranes in the cell (16–20). The members of the adaptor protein (AP) family, AP1–AP5, are tetrameric complexes consisting of two large and two small subunits. A second family of adaptor proteins is formed by the clathrin-associated sorting proteins (CLASP), a collection of monomeric proteins including AP180, epsin, and Eps15 (20,21). The global structure of the members of both families is very similar: they consist of a neatly folded section that binds to the membrane and a long disordered segment with clathrin binding motifs. Members of the AP-family possess a second long disordered segment, to attract assisting proteins. Of all adaptor proteins (henceforth abbreviated as "AP", irrespective of family), probably the most studied adaptor protein is the AP2 complex regulating endocytosis, which will also be the reference point in this study (11,22–24).

In addition to linking clathrin to membrane and cargo, a main function of APs is to regulate the assembly of clathrin cages by binding to multiple triskelia simultaneously. A series of in vitro experiments established that clathrin proteins in solution can be induced to self-assemble by adding APs (10,16). Recent structural studies revealed that AP2 can adopt two configurations, i.e., a closed state with part of the linker blocked from interacting with clathrin, and an open state where AP2 can bind two triskelia (25,26). With AP2 adapting the open state only when bound to a membrane, the formation of clathrin cages in a cell is effectively limited to the membrane. This mechanism may also explain why the in vitro assembly behavior of clathrin varies with the preparation state of the adaptor proteins, with well-cleaned adaptors inducing less activity (27). Our objective in this study is this little-explored question: Beyond the ability to bind two triskelia simultaneously, what else is required of APs to induce the formation of clathrin cages in solution?

The presence of an AP binding site at the end of each clathrin leg, a location henceforth informally referred to as the "toes" by following the common analogy of the clathrin leg with the human leg (see Fig. 1), is well established. Experiments with recombinant clathrin fragments indicate that

*Submitted December 14, 2015, and accepted for publication June 1, 2016.*
*Correspondence: w.k.denotter@utwente.nl
Editor: Markus Deserno.
http://dx.doi.org/10.1016/j.bpj.2016.06.003

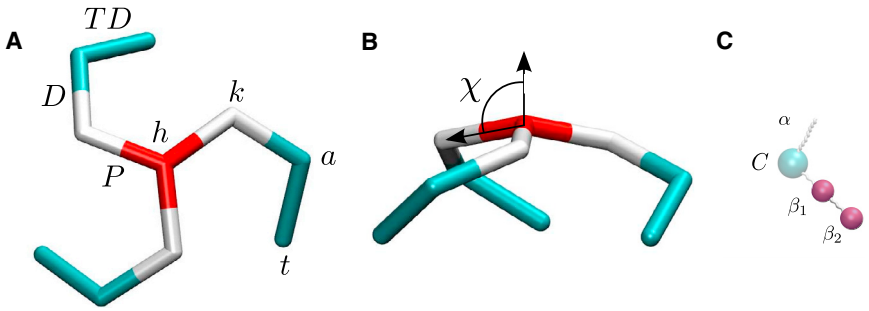


FIGURE 1 The highly coarse-grained simulation models of (*A* and *B*) clathrin and (*C*) AP2 on the same scale. In the rigid clathrin model, three proximal leg segments (*P*) radiate from a central hip (*h*) to the knees (*k*), at a pucker angle $\chi$ relative to the symmetry axis, followed by distal leg segments (*D*) running to ankles (*a*) and terminal domains (TD) ending at the toes (*t*). The AP model features two binding sites for clathrin, $\beta_1$ and $\beta_2$, connected by a flexible linker. In the full AP2 protein, the $\beta$-linker connects to a folded core (*c*) and a flexible $\alpha$-linker; these are omitted in the simulations because they do not play a role in the in vitro assembly process. To see this figure in color, go online.

this binding site is crucial to the inducement by AP2 of cage formation (28). At least one additional binding site, also required for cage formation, resides higher up each leg. Experiments with clipped triskelia point at a location on the trimer hub (29), i.e., in the region extending from the "hip" to just beyond the "knee" (see Fig. 1). Pull-down experiments identified a binding site near the "ankle" (30). Both options will be explored here.

Besides in vivo and in vitro experiments, the assembly behavior of clathrin has also been explored by in silico studies. In earlier work, two of us developed a highly coarse-grained patchy particle model of clathrin as a rigid triskelion with either straight or bend legs, and showed that anisotropic leg-leg interactions are the key to self-assembly (31,32). Simulations with this model that predicted a binding energy of ~23 $k_BT$ per clathrin in a cage, suggested a novel scenario, to our knowledge, for the transition from flat plaque to curved coat and yielded an assembly timescale in reasonable agreement with experiments (33,34). Matthews and Likos (35) modeled clathrin as a collection of 13-bead patchy particles, endowed with anisotropic interactions, and showed how these triskelia deformed a lipid membrane into a bud. Cordella et al. (36) and VanDersarl et al. (37) modeled clathrin as a spherical particle with anisotropic interactions accounting for three straight legs, and studied, among other properties, how a membrane influences an adjacent clathrin lattice. Adaptor proteins, which are crucial in bringing triskelia together under in vivo conditions, have been omitted in all clathrin simulations to date.

To address our research question, we apply coarse-grained simulations and statistical-mechanical theory to explore the ability of APs to induce the assembly of triskelia cages in solution. Because the AP model is based on the aforementioned key features, it is to be expected that other adaptor proteins can be modeled in a similar way. This article is organized as follows. In Materials and Methods, the clathrin simulation model is briefly discussed, the matching AP simulation model is introduced, and the implementation of click-interactions in Monte Carlo simulations is described. The findings on simulations of mixtures of triskelia and APs are presented and interpreted in Results and Discussion. The deduced qualitative understanding is then translated into a fairly simple quantitative theory, obtaining remarkably good agreement with simulations and experiments. We end with Conclusions.

## MATERIALS AND METHODS

In several preceding studies (31–34), we modeled clathrin as rigid patchy particles with three identical curved legs (see Fig. 1). The three legs are connected at a central hub, at a pucker angle $\chi$ relative to the threefold rotational symmetry axis of the particle, reflecting clathrin's intrinsic nonzero curvature. We here select a pucker angle, $\chi = 101^\circ$, typical of soccer-ball cages containing 60 triskelia, which is the most common cage size for in vitro experiments in the presence of AP (38). Each leg consists of two segments (i.e., the proximal and distal sections; the terminal domains were not included because of their expected small contribution to the clathrin-clathrin binding interaction) connected at the knee under a fixed angle and ending at the ankle. All leg segments are straight and of identical length, $\sigma = 17$ nm. The orientation of the distal segments relative to the proximal segments was chosen to allow maximum overlap between a particle and a secondary particle whose hub is situated at a knee of the primary particle. In a completed cage, a hub is located at every vertex—on top of three knees and three ankles of neighboring and next-nearest triskelia, respectively. A lattice edge is thus composed of two proximal and two distal segments, where the amino-acid sequences in both pairs of like segments run in opposite directions (i.e., anti-parallel). The attractive interaction between any pair of segments, which for clathrin is believed to result from a multitude of weak interaction sites along the legs (39–41), is modeled by a four-site potential based on the distances between the end-points of the two segments, with a minimum value of $-\epsilon$ for two perfectly aligned segments, as described in detail in the Supporting Material. The interaction is anisotropic under rotations around the long axes of the segments, to reflect that the binding sites are most likely concentrated on one side of the segment, to wit, the side that in a cage edge faces the three adjacent segments. Simulations revealed that this anisotropy of the attractive potential is crucial for the spontaneous self-assembly of triskelia into polyhedral cages (31,32). Excluded volume interactions between triskelia were omitted for computational reasons: this requires a more complex particle shape with nonlinear proximal and distal segments, as well as demands some flexibility of the legs, for the particles to pack together into cages with four legs interweaving along each edge, while the simulation step has to be reduced to prevent the relatively thin legs from crossing each other. Excluded volume interactions are important to prevent triskelia from binding to a cage edge in a slot that is already occupied by another triskelion; this property is incorporated into the simulation model by a repulsive potential between parallel segments of the same type. The moderate flexibility of the clathrin protein extends its interaction range beyond that of a rigidified protein; this effect is to some extent accounted for by the enlarged range of the intersegmental potential. The terminal domains (TDs) at the ends of the legs (see Fig. 1) were not included in our previous simulations, but they are required in

this study as binding sites for APs. The length and orientation of the TDs with respect to the proximal and distal segments were estimated using the structural information file PDB: 1XI4 for a clathrin cage (39), available at the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). Because the TDs are approximately equally as long as the proximal and distal segments, they are all assigned the same length $\sigma$ in the model. The TDs are attached to the ankle at an angle of 114° relative to the distal domain, with the three segments of a leg forming a dihedral angle of 28°. The clathrin-clathrin interactions are kept identical to those in the previous model; the TDs do not contribute to these interactions.

Continuing in this reductionist approach, we here introduce a matching simulation model of an AP (see Fig. 1). The model comprises the part of the AP2 protein that is involved in clathrin binding, i.e., the C-terminal region of the $\beta$-linker comprising the clathrin-box LLNLD of residues 631–635, the clathrin-binding appendage domain formed by residues 705–937, and the flexible linker connecting these two interaction sites (22). Our coarse-grained representation of this AP2 fragment consists of two point particles, embodying the two binding sites, connected by a tether. Because the remainder of the AP2 tetramer does not partake in clathrin binding and assuming that AP2s do not bind to each other, the omission of the majority of the protein is of no further consequence to the cage assembly process studied here. Excluded volume interactions are again omitted for reasons of computational efficiency; we note that the interior volume of a cage is far larger than the collective volume of the APs bound to a cage. The short range of the clathrin-AP binding interaction is inconvenient from a numerical point of view (see below). Instead, we developed a potential in which the $\alpha$th binding site on the $i$th triskelion and the $\beta$th particle of the $j$th AP dimer are bound with a fixed energy $-\zeta$ and are limited to a maximum separation $\rho$ in the clicked state ($b_{i\alpha,j\beta}=1$), while there are no interactions between these sites in the unclicked state ($b_{i\alpha,j\beta}=0$). As a function of the distance $r_{i\alpha,j\beta}$, the interaction potential then reads as

$$\phi_{\text{click}}\left(r_{i\alpha,j\beta},b_{i\alpha,j\beta}\right)=\begin{cases}0 & \text{for} \quad b_{i\alpha,j\beta}=0\\ \begin{cases}-\zeta & \text{for} \quad r_{i\alpha,j\beta}<\rho\\ \infty & \text{for} \quad r_{i\alpha,j\beta}\geq\rho\end{cases} & \text{for} \quad b_{i\alpha,j\beta}=1,\end{cases} \tag{1}$$

where $\zeta>0$, as illustrated in Fig. S4 in the Supporting Material. Because excluded volume interactions between AP2 tetramers ensure that a binding site on a clathrin can host at most one AP site, the clicks in the simulation model are constructed to be mutually exclusive: a site can partake in one click only. The clicks are also specific: the $\beta_1$ AP bead solely binds to the end of the TDs, i.e., at the toes, while the $\beta_2$ bead clicks only to a site higher up a triskelion's leg.

The two clathrin binding sites of AP2 are connected by an essentially structureless sequence of ~70 residues (22). According to polymer theory, this flexible linker will effectively act as an entropic spring with a spring constant $k$ and a maximum length $L$ (42,43). This behavior is modeled here by the finite extensible nonlinear elastic potential (44),

$$\phi_{\text{linker}}\left(l_j\right)=\begin{cases}-\frac{1}{2}kL^2\ln\left[1-\left(l_j/L\right)^2\right] & \text{for} \quad l_j<L\\ \infty & \text{for} \quad l_j\geq L,\end{cases} \tag{2}$$

where $l_j$ denotes the length of the $j$th AP dimer. The spring constant of an entropic spring is given by (43)

$$k=\frac{3k_{\text{B}}T}{2Ll_p}, \tag{3}$$

where $l_p$ is the persistence length. Given an average residue length of 0.37 nm, the linker of 70 residues connecting the two clathrin binding sites has a contour length of $L\approx 26$ nm $\approx 1.5\sigma$. Combination with the experimental value $l_p\approx 0.6$ nm for disordered proteins then yields $k\approx 30k_{\text{B}}T/\sigma^2$ for the linker.

The assembly characteristics of the combined models were simulated by the Monte Carlo (MC) method, i.e., by the weighted acceptance of randomly generated changes of the system configuration (45–47). Suppose that, by a sequence of steps, the system arrives in state $m$. In the MC technique, the transition probability from this state $m$ to a new state $n$ is expressed as

$$P_{m\to n}=P^{\text{trial}}_{m\to n}P^{\text{acc}}_{m\to n}, \tag{4}$$

where $P^{\text{trial}}_{m\to n}$ denotes the probability of generating the trial configuration $n$ from state $m$, and $P^{\text{acc}}_{m\to n}$ is the probability of accepting $n$ as the next state in the sequence of states; if the move is rejected, the system remains in the old state and $m$ is added (again) to the sequence of sampled states. For a symmetric trial move generator, $P^{\text{trial}}_{m\to n}=P^{\text{trial}}_{n\to m}$, the acceptance probability,

$$P^{\text{acc}}_{m\to n}=\min(1,\exp\{-\beta[\Phi(m)-\Phi(n)]\}), \tag{5}$$

where $\Phi(m)$ denotes the potential energy of state $m$ and $\beta=1/(k_{\text{B}}T)$, will produce a sequence of states in agreement with the equilibrium Boltzmann distribution.

The algorithm employed in this study applies two different types of trial moves, namely trial moves that alter the positions and orientations of particles, and trial moves that alter the connectivity between particles. The type of move is selected at random in every MC step, with positional moves selected $f$ times as often as connectivity moves. Positional trial moves start by randomly selecting a protein. If a clathrin is selected, its center of mass is displaced along all three Cartesian directions by random values in the range $[-(1/4)\sigma,(1/4)\sigma]$, and the particle is rotated around a random axis through the center of mass over a random angle in the range $[-(1/2),(1/2)]$ rad. A known complication in MC simulations is the drastic reduction of the mobility of particles interacting with neighbors, relative to the mobility of noninteracting particles, as can be seen clearly in movies of MC simulations (32). This is a minor issue in the assembly of cages from a solution containing clathrin only, as the free triskelia readily diffuse to a nearly immobile cage fragment. In simulations of mixtures of clathrin and AP, however, the binding of APs to triskelia will slow down their combined diffusion and hence significantly delay their attachment to cage fragments, especially if the AP-clathrin bond is strong and short-ranged. The solution adopted here is to apply cluster moves (45,48), i.e., the AP beads clicked to the selected triskelion move together with this clathrin, maintaining the statuses $b_{i\alpha,j\beta}$ and distances $r_{i\alpha,j\beta}$ of all clicks. Consider an AP with a bead clicked to the selected triskelion. If its other bead is unclicked or clicked to the same triskelion, the entire AP is moved with the clathrin as if they formed a rigid unit. If the AP's other bead is clicked to another clathrin, then this second bead is excluded from the trial move and, consequently, the length of the AP changes in the trial move. Next, the move is accepted or rejected following Eq. 5. If in a positional trial move an AP is selected, its two beads will be displaced independently. An unclicked bead is displaced in all three Cartesian directions by random values in the range $[-(1/4)\sigma,(1/4)\sigma]$, while a clicked bead is moved to a random position within a sphere of radius $\rho$ centered around the clathrin's matching clicking site. Next, the move is accepted or rejected following Eq. 5. Again, the statuses $b_{i\alpha,j\beta}$ of all clicks are conserved by these trial moves. In a clicking trial move, an AP bead is selected at random. The neighborhood of radius $\rho$ around this particle is scanned for matching clicking sites on triskelia; for a bead that is already clicked, its current partner will be among the $K$ detected sites. The unclicked state is included as the zeroth option. Instead of the above selection and acceptance steps, we directly accept one of the $K+1$ trial states as the next state. The probability of selecting the $k$th option is given by

$$P_k=\frac{\exp\left(-\Delta\phi^{\text{click}}_k\right)}{\sum_{k\prime=0}^{K}\exp\left(-\beta\Delta\phi^{\text{click}}_{k\prime}\right)}, \tag{6}$$

where the energy change $\Delta\phi_k^{\text{click}}$ between the old state and the $k$th trial state can only yield the values $\Delta\phi_k^{\text{click}} = \zeta$ for an unclicking trial move; $\Delta\phi_k^{\text{click}} = -\zeta$ for a clicking trial move; and $\Delta\phi_k^{\text{click}} = 0$ if the connection remains (un) clicked.

A number of simulations were run to verify that the unconventional click-potential and click-dependent MC cluster-moves sample the correct equilibrium distribution. Simulations with 1000 clathrins and 3000 APs in a cubic box of volume $10^6\sigma^3$ were used to determine the equilibrium constants of the reactions between triskelia and AP, defined as

$$K_{n,m}^{\text{tri}} = \frac{[CA'_nA''_m]_0}{[C]_0[A]_0^{n+m}}, \tag{7}$$

where $[C]_0$, $[A]_0$, and $[CA'_nA''_m]_0$ denote, respectively, the concentrations of unbound triskelia, unbound APs, and triskelia complexes with $n$ single-bound and $m$ double-bound adaptor proteins, in molars (see Appendix I). To improve the sampling efficiency, we reduced the number of distinct reaction products to three by reducing the number of clicking sites per triskelion from six to two—at the toes and ankle of the same leg—and reducing the entropic spring constant to $k = 1\ k_BT/\sigma^2$, while retaining the maximum extensibility of $1.5\sigma$. Furthermore, to enable comparison with exact analytical solutions, the adaptor proteins were not allowed to click to two clathrin particles simultaneously and the interactions between triskelia were turned off. Fig. S5 shows the equilibrium constants for triskelia that click once with an AP, $K_{1,0}^{\text{tri}}$, and for triskelia that bind two APs, $K_{2,0}^{\text{tri}}$, as functions of the clicking energy. Excellent quantitative agreement is observed with the statistical mechanical reaction equilibrium theory presented in Appendix I, which is shown in the graph as straight lines. Additional simulations confirm that equilibrium constants scale with the clicking radius $\rho$ conform to the power-law dependence derived in Appendix I (data not shown). The graph also shows the equilibrium constants for APs that double-click to a clathrin leg, $K_{0,1}^{\text{tri}}$, i.e., both sites of the AP are bound to the same triskelion leg. This occurs because the estimated maximum extensibility of the AP linker, $L \approx 26$ nm, well exceeds the length of the TD, $\sigma \approx 17$ nm, although the considerable elongation of the AP linker makes this double-click unfavorable. Again, the equilibrium constant is in good agreement with the theory. Several simulations were run with smaller systems to verify that the translation-versus-click-attempt ratio does not affect the results presented in this article; we settled on a value of $f = 10$ for reasons of computational efficiency.

The production simulations were all run with 1000 triskelia confined to a cubic box of volume $10^6\sigma^3$ with periodic boundary conditions. The number density of one triskelion per $10^3\sigma^3$ corresponds to an in vitro condition of ~0.2 mg/mL. Self-assembly in the absence of APs is observed in vitro for a slightly acidic solution (pH 6.2, 20 mM $MgCl_2$), with a critical assembly concentration (CAC) of ~0.1 mg/mL (49), i.e., the overall concentration where the fractions of bound and unbound triskelia are equal. In an earlier simulation study, we established that this concentration is the CAC of coarse-grained triskelia that gain $E_c \approx 23\ k_BT$ upon binding to a cage, which is realized for a segment-segment interaction parameter $\epsilon \approx 6\ k_BT$ (33). There we also showed that concepts borrowed from the thermodynamics of micelles allow a theoretic derivation of the binding energy from the measured CAC. Muthukumar and Nossal (50) extended these ideas with energetic contributions reflecting the curvature of the clathrin coat and applied them to analyze cages grown in the presence of AP2, even though the adaptor molecules themselves were not included in the theoretical model. A novel, to our knowledge, statistical mechanical derivation linking the binding energy to the CAC, by considering a cage as a collection of $p$ rigid triskelia with highly restricted translational and rotational freedom, is presented in Appendix II. For the assembly reaction $pC \rightleftharpoons C_p$, we obtain a standard state free difference of

$$\Delta G_p^0 = \mu_{C_p}^0 - p\mu_C^0 \approx p\Delta\mu_C^0, \tag{8}$$

with $\mu_X^0$ as the standard reference chemical potential of component $X$ and $\Delta\mu_C^0 \approx -16.4\ k_BT$ deduced from the CAC. Applied to the simulation model, this translates into a binding energy $E_c \approx 27\ k_BT$, in good agreement with the simulations. Recent experiments on the mechanical properties of clathrin coats adjacent to membranes confirm the binding (free) energies predicted by simulations and theory (51).

## RESULTS AND DISCUSSION

### Simulations

The effect of model APs on the self-assembly behavior of model triskelia is studied by systematically varying the clathrin-clathrin interaction $\varepsilon$, the AP-clathrin clicking strength $\zeta$, and the AP/clathrin ratio. Fig. 2 shows the assembly behavior on two cross sections of this three-dimensional parameter space, for the AP model clicking to the ankles and toes of clathrin. Every marker represents five independent simulations of $10^{10}$ MC steps, requiring approximately a week each on a desktop computer. Red crosses mark conditions where no spontaneous self-assembly of sizable cage fragments is observed. Green and blue circles indicate the self-assembly of at least one complete cage across the five simulations. For the green circles, $\epsilon > 6\ k_BT$, cages already self-assemble in the absence of APs. The blue circles highlight conditions where triskelia do not self-assemble in the absence of APs but do form cages in their presence—this is the region of parameter space where APs induce and control the formation of clathrin cages. The assembly of cages in the green and blue regions proceeds by a nucleation and growth process, just like in clathrin-only simulations

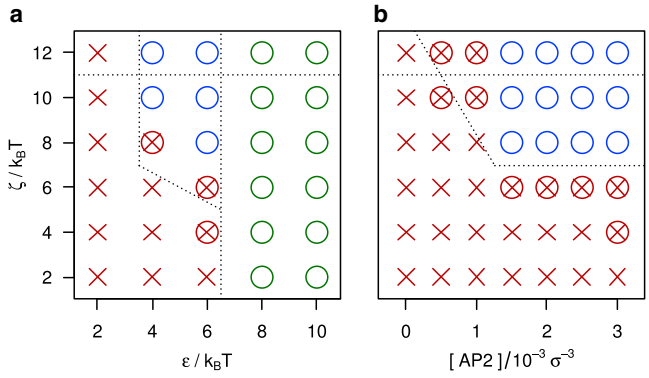


FIGURE 2 Cage assembly diagrams for clathrin, for 1000 triskelia at a concentration of $10^{-3}\ \sigma^{-3}$, combined with model APs clicking to the ends of the TDs and the ankles of clathrin, (*a*) as a function of the clathrin-clathrin binding strength $\varepsilon$ and the clathrin-AP clicking strength $\zeta$, at an AP/clathrin ratio of 3, and (*b*) as a function of the AP concentration (for AP/clathrin ratios from 0 to 3) and the clathrin-AP clicking strength, at a clathrin-clathrin binding strength of $\epsilon = 6\ k_BT$. The markers denote parameter combinations that result in the self-assembly of cages (a *green circle* if cages are also formed in the absence of APs; a *blue circle* if assembly only proceeds in the presence of APs), combinations that do not yield cages (*red crosses*), and conditions where cages do not assemble spontaneously but preassembled cages appear stable (*red cross* in *red circle*). The dashed lines indicate the approximate locations of phase boundaries, as discussed in more detail in the main text. To see this figure in color, go online.

(31,34). Small clusters of a few triskelia and APs (see Figs. 3 *A* and 4 *A*) are formed and destroyed continuously. Occasionally, one of these small aggregates crosses the nucleation barrier and grows into a cage, as illustrated by the snapshots in Fig. 3. Because of the rigidity of the clathrin model, these cages are all of approximately the same size, containing ~60 triskelia in near-spherical polyhedra with 12 pentagonal and ~20 hexagonal faces. The average cage diameter of ~4.5$\sigma$ (~75 nm) agrees with that for cages grown in vitro in the presence of APs (38), which motivated our choice of a 101° pucker angle. Cages grown in simulations with and without AP particles are of the same size. For in vitro experiments, however, a size difference is observed between cages grown with AP and cages grown without AP (38). It is unclear whether this difference is caused by the presence of APs, or by the pH reduction to induce cage formation in the absence of APs. We note that the cage size is very sensitive to the pucker; a decrease from 101° to 100° increases the average cage size by ~10 particles (31). Almost all self-assembled cages are complete, i.e., triskelion hubs reside at every vertex. Only rarely do one or two vertices of a nearly complete cage remain unoccupied, presumably because the remaining vacancies are less favorable binding sites than the occupied slots. The high prevalence of completed cages indicates that all vertices in these cages are of approximately equal binding affinity, which appears to confirm the "probable roads" hypothesis by Schein and Sands-Kidner (52). For low attachment rates at the edge of a growing fragment, particles binding in an unfavorable way have a high probability of being released again before the defect becomes permanently incorporated in the lattice through the attachment of subsequent particles.

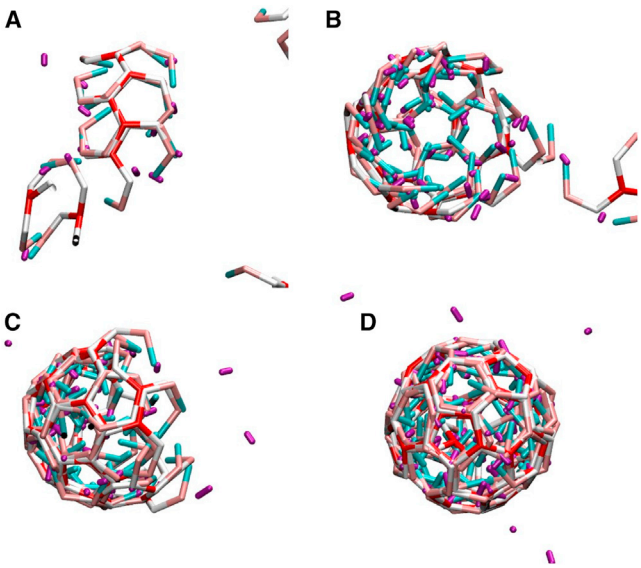


FIGURE 3 (*A–D*) A sequence of snapshots of triskelia assembling into a cage in the presence of APs, for $\epsilon = 6k_BT$, $\zeta = 8\,k_BT$, and an AP/clathrin ratio of 3, at intervals of $10^9$ MC steps. The coloring of the particles is the same as in Fig. 1. To see this figure in color, go online.

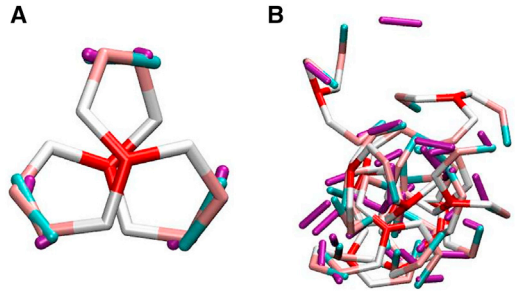


FIGURE 4 Adaptor proteins will bring triskelia together without regard for the relative positioning and orientation of these triskelia. A common aggregate (*A*) comprising two clathrins bonded by six APs (*purple*), saturating all clicking sites of the cluster. When the cluster is small and the interactions are weak, there are many opportunities to break the AP bonds and reshuffle the triskelia into a more favorable configuration. At high AP-clathrin clicking strengths, large disordered clusters develop rapidly (*B*); these will only very slowly acquire more order. To see this figure in color, go online.

Aggregation becomes frustrated when the binding energies are too strong. For intersegmental interactions exceeding ~10 $k_BT$, the triskelia easily stick together and thereby quickly form a multitude of small aggregates, which only very slowly merge into larger clusters. This evolution is reminiscent of that observed in vitro below pH 5.8 (53). A clicking energy exceeding ~11 $k_BT$ makes the APs eager to click to triskelia, thereby rapidly forming disordered clusters like that in Fig. 4 *B*, which only very slowly develop into cage fragments and ultimately, cages.

The rarity of nucleation necessitates excessively long simulations to accurately locate phase boundaries or to determine equilibrium cage concentrations (these will be obtained below by other means). The expedient used in the simulation phase diagrams of this section is the binary detection of self-assembled cages: green or blue circles if cages are formed, and red crosses otherwise. For phase points close to a phase boundary, additional simulations were initiated with configurations containing several half-spherical coats, to explore whether these aggregates grow into complete cages or disintegrate into monomers. In this context we note that the disassembly of an unstable coat fragment typically proceeds much faster than the completion of a stable fragment. The results of these simulations are included as green or blue circles or as red crosses in all simulation phase diagrams. For Fig. 2 only, a further refinement of the phase boundaries was obtained by running an additional set of simulations initiated with fully assembled cages stabilized by nearly three APs per clathrin (obtained from simulations at another phase point). The surviving cages are marked in Fig. 2 by red circles, superposed on the red cross indicating "no spontaneous assembly". If two simulations with the same parameter settings but opposite starting configurations converge to the same final state, it is very likely that this final state is the equilibrium state. If their final states differ, then either the stability

difference between these states is small or (at least) one of the simulations is trapped in a local minimum of the free energy landscape.

The dashed lines in Fig. 2 indicate the estimated phase boundaries, where the boundary slightly to the right of $\epsilon = 6\ k_BT$ was established previously and with greater accuracy (33) than the other boundaries. One sees in Fig. 2 *a* that, at the prevailing concentrations, the APs are able to regulate the emergence of cages for $4\ k_BT \lesssim \epsilon \lesssim 6\ k_BT$ and $\zeta \gtrsim 7\ k_BT$ (i.e., the *blue region*). A cross section of this region, by varying the AP concentration at fixed $\epsilon = 6\ k_BT$, is presented in Fig. 2 *b*. This plot shows that AP-induced cage assembly requires a clicking energy $\zeta \gtrsim 7\ k_BT$ as well as an AP concentration equal to or exceeding the clathrin concentration.

Besides the AP model discussed above, simulations were run with a number of alternative models to explore the conditions conducive to adaptor-induced cage formation. APs clicking at the knees and toes yield the assembly diagrams presented in Fig. 5. The graph on the left is similar to its counterpart in Fig. 2, and shows that APs binding at the knees are equally capable of regulating the assembly of cages as APs binding at the ankles. The graph on the right shows an interesting difference between the two cases: self-assembly continuous down to much smaller AP concentrations. Lowering the effective spring constant of the linker between the AP beads to $k = 10\ k_BT/\sigma^2$ has little impact on the assembly diagrams of either adaptor model (data not shown). Upon a further reduction to $k = 1\ k_BT/\sigma^2$ (see Fig. 6), the AP clicking to the knees and toes remains operational (with a slight shift in the smallest $\zeta$ inducing cage formation), while the AP clicking to the ankles and toes ceases to function.

To understand the results reported above, we now turn to unraveling the mechanism by which APs induce the aggregation of triskelia. The discussion presented here is qualitative in nature; a quantitative analysis of the insights gained is presented in the next section. Consider first the AP model

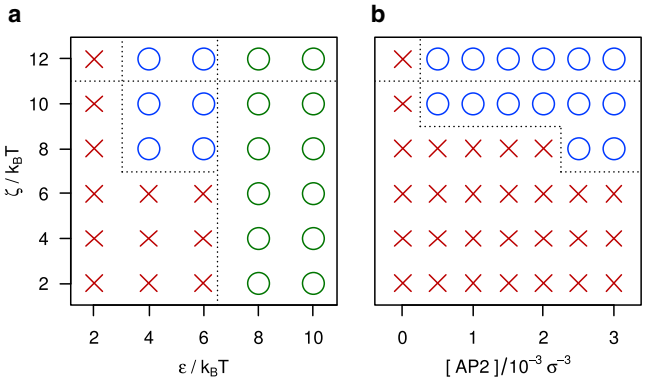


FIGURE 5 Assembly diagrams for model APs clicking to the ends of the TDs and the knees of clathrin, with all other conditions and markers in (*a*) and (*b*) identical to those in Fig. 2 *a* and *b*, respectively. The blue circles again highlight the parameter space where cage formation is controlled by APs. To see this figure in color, go online.

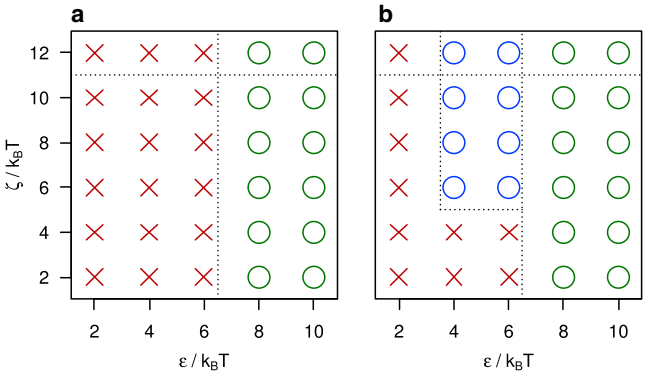


FIGURE 6 Assembly diagrams for model APs with a reduced (entropic) spring constant, $k = 1\epsilon/\sigma^2$; all other parameters are identical to those in Fig. 2 *a* and 5 *a*. APs clicking to the ankles and TDs of clathrin (*a*) are no longer able to regulate the formation of cages, while APs clicking to the knees and TDs of clathrin (*b*) are still operational. To see this figure in color, go online.

that binds to the toes and the knees. It is clearly energetically favorable for an AP to click to triskelia. The largest gain in energy is obtained when the adaptor clicks twice, which is only achieved—note that the toe-knee distance in a clathrin is longer than the maximum extensibility of the linker—if the AP binds to two distinct triskelia. Bringing two triskelia together strongly enhances their chances of adopting the correct relative positions and orientations, and hence promotes successful binding. Adaptor proteins may thus contribute to both the stability of clathrin aggregates and the rate at which they are formed. Note that this line of thought assumes that the energetic gain upon binding outweighs the accompanying entropic loss in translational freedom (and in rotational freedom for clathrin-clathrin binding) and thereby lowers the overall Helmholtz free energy of the system. Hence, whether the AP plays a supporting role in cage formation depends on the clicking strength as well as on the AP and clathrin concentrations.

For the adaptor clicking at the toes and ankle, the energetic gain upon double-clicking to one clathrin is identical to that of clicking to two triskelia. This partially invalidates the mechanism proposed above, by providing the APs with an alternative binding option that does not contribute toward cage assembly. Yet, the simulations of Fig. 2 indicate that these adaptors are able to induce cage formation. Inspection of the length distribution of the linkers (data not shown) reveals that 1) most APs bound to a cage are bridging between pairs of triskelia, and 2) the nearest toe-ankle distance in a cage is shorter than the toe-ankle distance of $1\sigma$ along a clathrin leg. This suggests that the shorter linker length in a cage, and between triskelia in the process of coming together, results in a lower elastic energy and hence a higher Boltzmann factor, and thereby favors APs connecting between sites on distinct triskelia over APs connecting to two sites on the same clathrin. The reader might note that

the distribution of end-to-end distances of the real linker is determined by entropic effects, while this distribution is modeled here as an energetic effect (see Eq. 2), but this does not present any conceptual problem as both yield the same dependence of the free-energy on the interbead distance.

In support of the above considerations, we recall the impact on the assembly behavior of reducing the linker spring constant at constant maximum extensibility (see Fig. 6). For the model AP clicking at toes and knees, the reduction of the spring constant was of little consequence, in agreement with the mechanism where an adaptor clicking twice always establishes a link between two distinct triskelia. For the model AP clicking to toes and ankles, however, lowering the spring constant reduces the difference in internal energy between AP double-clicked to one clathrin (with the linker stretched to $1\sigma$) and AP clicked to two triskelia (with a shorter linker length). With this reduction, the preference for interclathrin over intraclathrin bonds diminishes and, at $k = 1\ k_BT/\sigma^2$, the number of APs links holding triskelia together becomes too low to stabilize a cage.

## Theory

A statistical mechanical theory of AP-induced cage assembly, built on the concepts deduced above, is derived in Appendix III. The theory predicts the equilibrium constant $K^{\text{cage}}_{p,n,m}$ relating the concentrations of unbound triskelia and unbound APs to the concentration of cages of $p$ triskelia decorated with $n$ single-clicked APs and $m$ intertriskelion double-clicked APs. Suppose one knows the average binding energy of a triskelion in a cage devoid of APs, $E_c$; the clathrin-AP interaction strength $\zeta$; and the total concentrations of clathrin and AP in a system, $[C]_t$ and $[A]_t$, respectively. It is now possible to compute the equilibrium concentrations of all decorated cages in that system, $[C_pA'_nA''_m]$, by the iterative procedure outlined in Appendix III; the overall cage concentration then follows by a summation over all decorated cages, i.e., all values of $p$, $n$, and $m$. Because the simulations predominantly produced cages of 60 triskelia, we restrict the theoretical calculations to one cage size, $p = 60$. The phase diagrams calculated for the ankle-binding AP model are shown in Fig. 7. To facilitate the comparison with the simulation results in Fig. 2, the plots are based on the same total clathrin concentration, $[C]_t = 10^{-3}\sigma^{-3}$, and similar interclathrin binding energies. In theory, the maximum binding energy due to interclathrin interactions amounts to $E_c = 6\epsilon$ per triskelion in a cage. In practice, due to thermal vibrations and the inevitable alignment mismatches in cages formed by rigid identical particles, the average potential energy in the simulations is given by $E_c \approx 4\epsilon$ (33). The latter relation has been used to rescale the horizontal axes of several phase diagrams in this section for ease of comparison with simulation results. For increasing binding strengths at constant AP concentration, Fig. 7 *a* shows a narrow transition region (*yellow*) between virtually no cage formation (*dark red*) and almost all triskelia absorbed in cages (*dark green*). A more gradual transition with increasing AP concentration is observed in Fig. 7 *b*. Considering the relative simplicity of the theory, the good agreement between Figs. 2 and 7 is very satisfactory. The theory does not reproduce two properties observed in the simulations: there are no disordered aggregates at high clicking energy, because this transient intermediate state is not included in the theory, and the self-assembly for $\zeta \gtrsim 10\ k_BT$ continues down to low AP concentrations. The latter confirms our earlier suspicions that the self-assembly simulations have not reached equilibrium, and agrees with the observation that preassembled cages appear

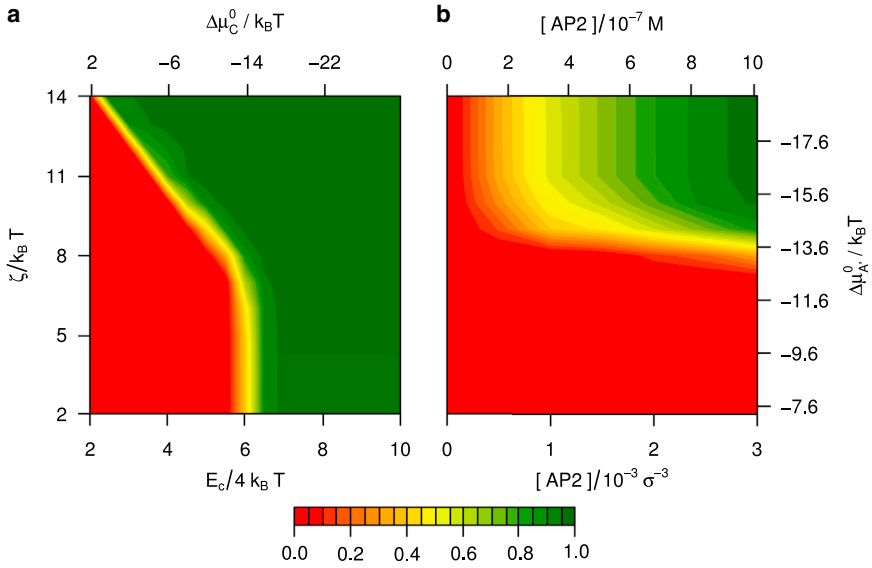


FIGURE 7 Assembly diagrams calculated using the theory derived in Appendix III, showing the fraction of clathrin bound in cages, for APs clicking to the ends of the TDs and the ankles of triskelia: (*a*) as a function of the binding energy per clathrin in an AP-free cage, $E_c$, and the clathrin-AP clicking strength, $\zeta$, at an AP/clathrin ratio of 3; and (*b*) as a function of the AP concentration (for AP/clathrin ratios from 0 to 3) and the clathrin-AP clicking strength, at $E_c = 22\ k_BT$. The two graphs refer to the same total clathrin concentration, $[C]_t = 10^{-3}\sigma^{-3} \approx 3.4 \times 10^{-7}$M, and similar interaction energies, as their counterparts in Fig. 2. For comparison purposes, the horizontal axis of the left plot is scaled by the simulation-based ratio $E_c/\epsilon \approx 4$ (see main text). The alternative axes to the graphs are labeled with the standard chemical free energy differences of AP single-clicking to clathrin (see Eq. 23), and of clathrin assembling into AP-free cages (see Eq. 33), and with total AP concentrations in molars. To see this figure in color, go online.

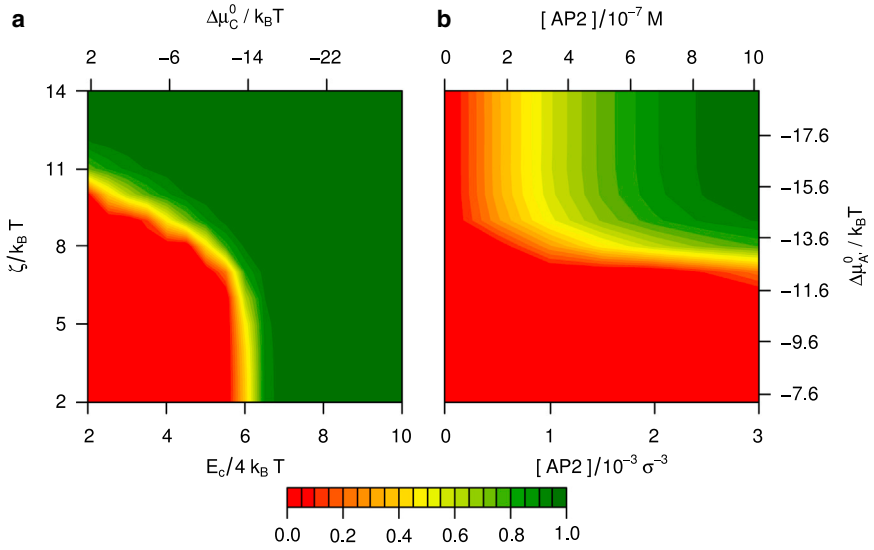


FIGURE 8 Calculated fraction of clathrin bound in cages, for APs clicking to the ends of the TDs and the knees of triskelia, with all other conditions in (*a*) and (*b*) equal to those in Fig. 7 *a* and *b*, respectively. These graphs are the theoretical counterparts to the simulation results in Fig. 5 *a* and *b*, respectively. To see this figure in color, go online.

stable under these conditions (see the *red crossed circles* in the *top-left* of Fig. 2 *b*).

Calculated phase diagrams for the AP model binding to knee and toes are presented in Fig. 8, and compare well with the diagram deduced from the simulations (see Fig. 5). The striking resemblance between the calculated phase diagrams (compare Figs. 7 and 8) suggests that the sole difference between the two calculations, i.e., an AP model that can double-click to a single clathrin versus an AP model that cannot, is of little consequence to the equilibrium behavior. The main difference, the slope of the yellow phase boundary in the plots on the left, results from APs double-clicking to triskelia. For APs binding to the knee, intraclathrin double-clicks are impossible. Double-clicks are unlikely at moderate click strengths for APs binding to the ankle, because of the free energy penalty in stretching the AP linker, but they become important at high click strengths. The phase diagrams calculated for a reduced linker spring constant of $k = 1\epsilon/\sigma^2$ also agree well with the simulations: the model APs binding to the ankle do not induce cage assembly, while the model APs binding to the knee continue to function (data not shown). Collectively, these results provide strong support for the theory and the underlying concepts on the mechanism of cage stabilization by APs.

Under experimental conditions, the binding strengths $E_c$ and $\zeta$ are typically unknown constants, whose values are co-determined by the acidity and salt conditions of the solvent, while the concentrations are readily varied. Four assembly phase diagrams pertaining to various binding strengths are presented in Fig. S6. To facilitate comparison with experiments, the data are presented in terms of the standard chemical potential difference of AP single-clicking to clathrin, $\Delta\mu_{A'}^0$, as defined in Eq. 23, and the standard chemical potential difference of the formation of AP-free cages, $\Delta\mu_C^0$, as defined in Eq. 33. At $\Delta\mu_C^0 = -13.8\ k_BT$ (see Fig. S6 *a*), the triskelia readily aggregate in the absence of APs at the higher end of the clathrin concentration range; adding APs with $\Delta\mu_{A'}^0 = -15.3\ k_BT$ enhances the cage concentration, but the effect quickly saturates. For the slightly weaker binding triskelia at $\Delta\mu_C^0 = -11.8\ k_BT$, the assistance of APs is crucial to cage formation, with APs binding at $\Delta\mu_{A'}^0 = -14.3\ k_BT$ yielding significantly more cages than APs clicking at $\Delta\mu_{A'}^0 = -13.3\ k_BT$ (compare Fig. S6 *b* and *c*). An interesting feature is observed at even weaker clathrin bounding, $\Delta\mu_C^0 = -7.8\ k_BT$, in combination with $\Delta\mu_{A'}^0 = -15.3\ k_BT$ (see Fig. S6 *d*), where for a constant overall clathrin concentration, of say, $1.7 \times 10^{-7}$ M, the concentration of cages at first increases with the overall AP concentration, passes through a maximum, and then decreases with increasing AP concentration. This cross section is highlighted in Fig. 9, along with three profiles at lower and higher clathrin concentrations. A similarly shaped profile was obtained by the in vitro assembly experiments of Zaremba and Keen (38), but there the assembled protein mass fraction is plotted; curves of this type are also included in Fig. 9. These authors explain the local maximum as a saturation effect, with clathrin becoming the limiting component upon increasing the AP concentration. This effect is visible in the curves for $[C]_t = 3.3 \times 10^{-7}$M, which saturates in the fraction of bound clathrin but decays in the fraction of bound protein. Our calculations provide an additional explanation for a maximum in the assembled fraction: the number of cages decreases beyond an optimum AP concentration. The underlying mechanism is the replacement of double-clicked APs with two single-clicked APs each, thereby weakening the integrity of cages. Hence increasing the AP concentration beyond its optimum results in a reduction of the cage concentration, as can be clearly seen for $[C]_t = 1.7 \times 10^{-7}$M.

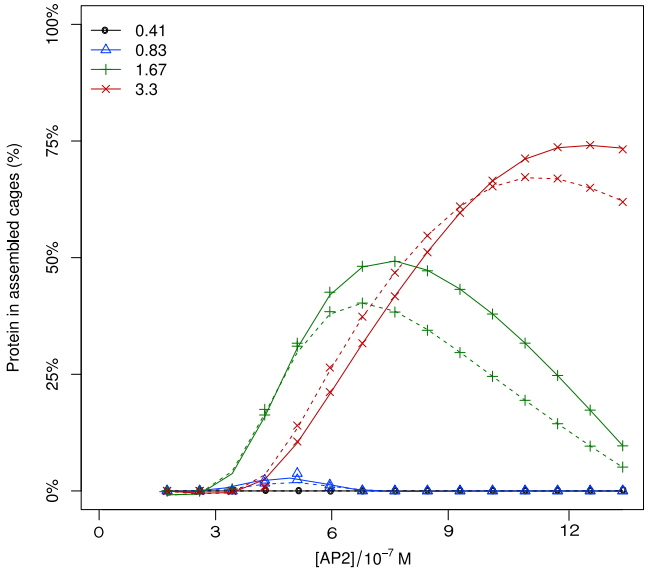


FIGURE 9 Calculated number fraction of clathrin (*solid lines*) and weight fraction of protein (*dashed lines*) in self-assembled cages, as a function of the total AP concentration. The total clathrin concentration is indicated in the legend, in units of $10^{-7}$ molar; the APs bind to the ankle and TD of clathrin, $\Delta\mu_C^0 = -7.8\ k_BT$ and $\Delta\mu_{A'}^0 = -15.3\ k_BT$. Note that the fractions bound in cages do not increase monotonically but pass through a maximum, for reasons explained in the main text. To see this figure in color, go online.

Plots of the number of APs bound to cages, normalized by the number of triskelia in a cage, are presented in Fig. 10. Because the cages are nearly saturated with double-clicked APs for the phase point explored in Fig. 9, we opted to present results for the chemical potential difference combinations in Fig. S6, *b* and *c*. Markers are plotted for cage concentrations exceeding $3 \times 10^{-10}$ M, which corresponds to one cage in the simulated system. At this threshold, the average number of double-clicked APs per encaged clathrin

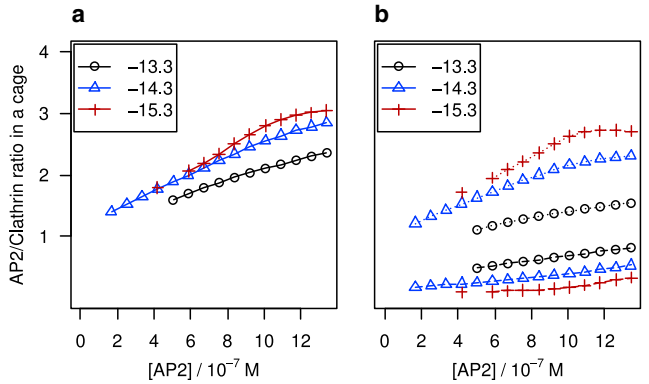


FIGURE 10 Calculated average AP/clathrin ratio for cages (*a*) and subdivision (*b*) into single-clicked (*dashed lines*) and double-clicked (*solid lines*), as functions of the total AP concentration. The clicking standard chemical free energy difference is indicated in the legend, in units of $k_BT$, the APs bind to the ankle and TD of clathrin, $\Delta\mu_C^0 = -11.8\ k_BT$ and $[C]_t = 10^{-3}\sigma^{-3}$. To see this figure in color, go online.

equals approximately one, i.e., a clathrin is clicked to two cross-linking APs on average, while the average number of single-clicked APs is substantially lower. With increasing AP concentration, the number of double-clicked APs rises with approximately the same slope as the number of single-clicked APs for $\Delta\mu_{A'}^0 = -13.3\ k_BT$, while for higher click strengths the number of double-clicked increases more than the number of single-clicked. In addition to the growing number of APs per cage, the number of cages also rises over the range of AP concentrations. For $\Delta\mu_{A'}^0 = -15.3\ k_BT$, the number of single-clicked only starts to deviate from zero when the number of double-clicked APs levels off, at ~2.7 AP per triskelion. These turning points coincide with the number of cages leveling off to a broad maximum, akin to those in Fig. 9.

## CONCLUSIONS

A coarse-grained simulation model and a theory were developed to study the AP-induced self-assembly of triskelia into cages. The results of both approaches are in line with the experimental data, and provide a better understanding of how APs regulate the assembly of cages. This study reveals a number of restrictions on functional APs. Clearly, APs must bind clathrin in a manner sufficiently strong to bring two triskelia together, but cage formation is frustrated when APs bind too strongly. The flexible linker between the two binding sites of an AP must be long enough for intertriskelion connections in cages, but the linker should not be too long to avoid intratriskelion bonding. On a related note, the effective spring constant of the linker must be weak enough to allow intertriskelion connections in cages, but not too weak to suppress intratriskelion bonding. And the AP/clathrin ratio must be high enough, although not too high. While the numerical values used in the model and theory are based on AP2, we expect the results to apply to all types of APs. For the advancement of simulation models and theories, as well as for an improved understanding of the thermodynamics of coat and vesicle formation during endocytosis, it would be useful to obtain experimental values of all binding constants involved, as well as of the mechanical properties of the AP linker. One way of measuring these parameters is proposed in Appendix III.

## APPENDIX I

### Clathrin-AP complexes

In these Appendices, expressions are derived for the reaction equilibrium constants of AP binding to a triskelion, clathrin self-assembly into cages, and AP-induced cage assembly, respectively. We start by considering a mixture of clathrin ($C$) and adaptor ($A$) proteins in equilibrium with their supramolecular aggregates by reactions of the type

$$C + (n+m)A \rightleftharpoons CA'_nA''_m, \tag{9}$$

where the primes represent the number of clicks binding an AP to the clathrin, in this case $n$ single-clicked and $m$ double-clicked APs. Because clathrin has six binding spots for AP, each capable of binding at most one AP, it follows that $n \geq 0$, $m \geq 0$, and $n + 2m \leq 6$. For simplicity, we assume these six sites to have identical binding properties. Likewise, the two clicking sites of AP are assumed to have identical properties, except for their specificity to either the TD or the ankle/knee binding site of clathrin. The equilibrium constant of the above reaction can be defined as (54)

$$K^{\mathrm{tri}}_{n,m} = \frac{\left([CA'_nA''_m]/c_0\right)}{([C]/c_0)([A]/c_0)^{n+m}}, \tag{10}$$

with the square brackets denoting concentrations, i.e., particles per unit of volume; and $c_0$ is a reference concentration typically taken to be 1 molar. From the statistical mechanics of reaction equilibria in ideal mixtures (54–56), it follows that

$$K^{\mathrm{tri}}_{n,m} = \frac{\left(q_{n,m}/V\right)}{(q_C/V)(q_A/V)^{n+m}} c_0^{n+m} = e^{-\beta\Delta G^0_{n,m}}, \tag{11}$$

where $q_C$, $q_A$, and $q_{n,m}$ denote the molecular partition functions of unbound clathrin, unbound AP, and the $CA'_nA''_m$ supramolecule, respectively; and $\Delta G^0_{n,m}$ is the standard state free-energy change of the reaction.

The semiclassical partition function of a rigid clathrin particle in an infinitely dilute solution, i.e., in the limit that nonbonded interactions can be ignored, is given by

$$q_C = \frac{1}{\Delta_C}\iint e^{-\beta\Phi} d\mathbf{r} d\boldsymbol{\varphi} \approx \frac{8\pi^2 V}{\Delta_C} e^{-\beta\Phi_C}, \tag{12}$$

with $\Phi$ as the interaction potential and $\Phi_C$ as the average solvation free energy of a clathrin. The position integrals run over the volume $V$ of the system and the three-dimensional orientation angles run over their entire range, e.g., for the Euler angles $\varphi_1 \in [0, 2\pi)$, $\varphi_2 \in [0, \pi)$, $\varphi_3 \in [0, 2\pi)$, and $d\boldsymbol{\varphi} = \sin\varphi_2 d\varphi_1 d\varphi_2 d\varphi_3$. The elementary volume element $\Delta_C$ follows from

$$\Delta_C^{-1} = \left(\frac{2\pi k_B T}{h^2}\right)^3 m_C^{3/2} |\mathbf{I}_C|^{1/2} \frac{1}{\sigma_C}, \tag{13}$$

with $h$ denoting Planck's constant; $m_C$ and $\mathbf{I}_C$ are the mass and inertia tensor of a triskelion, respectively; the brackets $|\ldots|$ denote a determinant; and where the symmetry number $\sigma_C$ has the value 3 for a particle with a threefold rotational axis. Note that $h$, $m_C$, and $\mathbf{I}_C$ do not enter the MC simulations, hence the theoretical and simulated equilibrium constants will only agree if these factors can be made to cancel out in the final expression. Treating an AP protein as two point particles of type $a$, one obtains at infinite dilution

$$q_A = \frac{1}{\Delta_a^2}\iint e^{-\beta\Phi} d\mathbf{r}_1 d\mathbf{r}_2 \approx \frac{1}{\Delta_a^2} V q_s\, e^{-\beta\Phi_A}, \tag{14}$$

where $\Delta_a = h^3/(2\pi m_a\, k_B T)^{3/2}$ is the elementary volume element per particle; $\Phi_A$ is the average solvation free energy of an AP; and

$$q_s = 4\pi \int_0^\infty e^{-\beta\psi(r_{12})} r_{12}^2 dr_{12}, \tag{15}$$

is the contribution of the internal spring, with potential energy $\psi(r_{12})$ at elongation $r_{12}$, to the partition function. The integral is readily solved for a Hookean spring with spring constant $k$, yielding $q_s = (2\pi\, k_B T/k)^{3/2}$.

Next, the partition function of a clathrin adorned with one single-clicked AP takes the form

$$\begin{aligned} q_{1,0} &= \frac{1}{\Delta_C \Delta_a^2}\int\int\int\int e^{-\beta\Phi} d\mathbf{r} d\boldsymbol{\varphi} d\mathbf{r}_1 d\mathbf{r}_2 \\ &\approx g_{1,0}\frac{8\pi^2 V}{\Delta_C}\frac{4\pi\rho^3 q_s}{3\Delta_a^2} e^{-\beta(\Phi_C+\Phi_A-\zeta)}, \end{aligned} \tag{16}$$

where, in the last step, it has been used that either site of the AP dimer must be within a sphere of radius $\rho$ centered around a clicking site on the triskelion, and $\zeta$ denotes the strength of the click. The number of clicking combinations will be denoted by $g_{n,m}$, and in this case has the value $g_{1,0} = 6$ because a triskelion offers six binding spots. Note that the AP-clathrin complex is not treated as a single molecule, but as a combination of two molecules with reduced rotational and translational freedom (57,58). By combining the above equations, one arrives at the equilibrium constant

$$K^{\mathrm{tri}}_{1,0} = g_{1,0}\frac{4}{3}\pi\rho^3 e^{\beta\zeta} c_0, \tag{17}$$

where the elementary volumes have indeed canceled out. The approach is readily extended to several single-clicked APs per triskelion, with at most one AP per triskelion binding site, under the assumption that other interactions between these APs may be ignored. Fig. S5 shows that the theory is in good agreement with the simulations.

The partition function of a triskelion adorned with one double-clicked AP is given by an integral similar to that in Eq. 16, with the restriction that now both sites of the AP must be clicked to their counterpart sites on the triskelion. In view of the estimated maximum extensibility of the AP linker, $L \approx 1.5\sigma$, a double-clicked AP will bind to two triskelion sites on the same leg and hence their interstitial distance is fixed, $d_t = 1\sigma$. We then arrive at

$$q_{0,1} \approx g_{0,1}\frac{8\pi^2 V}{\Delta_C}\frac{q''_s(d_t)}{\Delta_a^2} e^{-\beta(\Phi_C+\Phi_A-2\zeta)}, \tag{18}$$

where $g_{0,1} = 3$ and the contribution of AP's internal spring reads as

$$q''_s(d) = \int_{v_1}\int_{v_2} e^{-\beta\psi(|\mathbf{r}_1-\mathbf{r}_2|)} d\mathbf{r}_1 d\mathbf{r}_2, \tag{19}$$

with $v_1$ and $v_2$ denoting the spherical volumes of radius $\rho$ of two clicking sites at center-to-center distance $d$. For the actual proteins, the range of the click interaction is short compared to the distance between clicking sites, $\rho \ll d$, and the integral may be approximated as

$$q''_s(d) \approx \left(\frac{4}{3}\pi\rho^3\right)^2 e^{-\beta\psi(d)} \tag{20}$$

in the limit of $\beta k d\rho \ll 1$. Fig. S5 shows that the resulting equilibrium constant, $K^{\mathrm{tri}}_{0,1}$, is in good agreement with the simulations, for the low $k = 1\, k_B T/\sigma^2$ value used in that plot. The combination of spring constant $k = 30\, k_B T/\sigma^2$ and click radius $\rho = 0.25\sigma$ used in the production simulations exceeds this limit and it proved necessary to evaluate the integral of Eq. 19 numerically, yielding a value $q''_s(\sigma) \approx 9.6 \times 10^{-8}\sigma^6$ approximately two orders larger than the estimate $1.0 \times 10^{-11}\sigma^6$ by Eq. 20, to obtain a good agreement between theoretical and simulation phase diagrams.

The above results can be combined to obtain the equilibrium constants for all reactions of the type expressed in Eq. 9, in the dilute

limit. Upon neglecting interactions between APs bound to the same triskelion, except for the mutual exclusivity of the clathrin-AP clicks, we arrive at

$$K_{n,m}^{\mathrm{tri}} = g_{n,m}\left(\frac{4}{3}\pi\rho^3\right)^n\left(\frac{q_s''(d_t)}{q_s}\right)^m e^{\beta(n+2m)\zeta}c_0^{n+m}. \quad (21)$$

The multiplicity $g_{n,m}$ is readily established by counting the number of ways of attaching $n$ single-clicked and $m$ double-clicked APs to a single triskelion, but in practice this number proves of little consequence because the other factors in the above equation are much larger. Upon neglecting this factor, the standard state free energy differences (54,56) for the 15 possible $(n,m)$ combinations with $n+m>0$ reduce to

$$\Delta G_{n,m}^0 = \mu_{n,m}^0 - \mu_C^0 - (n+m)\mu_A^0 \quad (22)$$

$$\approx n\Delta\mu_{A'}^0 + m\Delta\mu_{A''}^0(d_t), \quad (23)$$

with $\mu_i^0$ as the reference chemical potential of compound $i$ at the reference concentration $c_0$, and where the reference chemical potential differences follow from Eq. 21 as

$$\Delta\mu_{A'}^0 = -k_{\mathrm{B}}T\ln\left(\frac{4}{3}\pi\rho^3 e^{\beta\zeta}c_0\right), \quad (24)$$

$$\Delta\mu_{A''}^0(d) = -k_{\mathrm{B}}T\ln\left(\frac{q_s''(d)}{q_s}e^{\beta 2\zeta}c_0\right). \quad (25)$$

Inserting the parameters of the simulation model into the former difference yields

$$\Delta\mu_{A'}^0 \approx -5.3\ k_{\mathrm{B}}T - \zeta, \quad (26)$$

while in combination with the approximation in Eq. 20, the latter difference can be rewritten as

$$\Delta\mu_{A''}^0(d) = 2\Delta\mu_{A'}^0 + \psi(d) + k_{\mathrm{B}}T\ln\left[c_0(2\pi k_{\mathrm{B}}T/k)^{3/2}\right], \quad (27)$$

and with the numerical evaluation of $q_{\mathrm{s}}''(d_{\mathrm{t}})$ we find for the simulation model

$$\Delta\mu_{A''}^0(d_{\mathrm{t}}) = 2\Delta\mu_{A'}^0 + 16.4\ k_{\mathrm{B}}T. \quad (28)$$

These expressions are readily extended to include site-dependent clicking strengths, i.e., $\zeta_1$ for binding to the feet and $\zeta_2$ for binding to the ankle or knee.

## APPENDIX II

## Clathrin cages

The partition function of a clathrin cage of $p$ triskelia is obtained by integrating over the positions $\mathbf{r}$ and orientations $\boldsymbol{\varphi}$ of all $p$ triskelia, subject to the condition that the particles remain sufficiently close and properly oriented—relative to each other—to qualify as a cage. The overall translational and rotational freedom of a triskelion—amounting to $V$ and $8\pi^2$, respectively, for a particle in solution (see Eq. 12)—are effectively reduced by these binding restrictions to $v_{\mathrm{t}}$ and $\omega_{\mathrm{r}}$, respectively, for a triskelion wobbling around a fixed location in a cage. The partition function of a cage of $p$ triskelia can therefore be approximated as

$$q_p = \frac{1}{p!\Delta_C^p}\iint\cdots\iint e^{-\beta\Phi}d\mathbf{r}_1 d\boldsymbol{\varphi}_1\cdots d\mathbf{r}_p d\boldsymbol{\varphi}_p \quad (29)$$

$$\approx g_p\frac{8\pi^2 V}{\Delta_{\mathrm{C}}^p}(v_t\omega_r)^{p-1}e^{-\beta p(\Phi_C - E_c)}, \quad (30)$$

where $E_c$ denotes the average binding energy of a clathrin in a cage, and $p!$ corrects for the indistinguishability of the $p$ triskelions forming the cage. In evaluating the integral, one particle has retained the full factor $8\pi^2 V$ to account for the translational and rotational freedom that a rigid coat will sample, while the remaining $(p-1)$ particles each contribute a factor $v_{\mathrm{t}}\omega_{\mathrm{r}}$ reflecting thermal fluctuations around this rigid shape. The multiplicity $g_p$ denotes the degeneracy of the ground state. Cages with pentagonal and hexagonal facets require an even $p$; there exists one cage structure for $p=20$, none for $p=22$, and multiple cage structures for $p\geq 24$. Schein and Sands-Kidner (52) and Schein et al. (59) argued that, for $20\leq p\leq 60$, there typically exists just one preferred cage structure for every even value of $p$, because all other cages incorporate one or more edges with an unfavorably high torsional energy. This theory is confirmed by the cages spontaneously grown in our simulations. We note that the "exclusion of head-to-tail dihedral angle discrepancies" rule proposed by Schein and Sands-Kidner (52) and Schein et al. (59) can be expressed much more concisely as the "excluded 5566" rule: an unfavorable torsion arises when a facet has among its neighboring facets a sequence of two pentagons followed by two hexagons, regardless of clockwise or counterclockwise order. The "isolated pentagon rule" applies for $p>60$, and there typically exist multiple favorable cages for $p\geq 70$ (52,59). Because the multiplicity is a small integer for the $p\approx 60$ cages grown in the simulations, the exact value of $g_p$ proves to be of little consequence to the results of the calculations.

Combining the above results, the equilibrium constant for the cage formation reaction

$$p\mathrm{C}\rightleftharpoons\mathrm{C}_p \quad (31)$$

is found to be given by

$$K_p^{\mathrm{cage}} = g_p\left(\frac{v_t\omega_r}{8\pi^2}\right)^{p-1}e^{\beta p E_c}c_0^{p-1}. \quad (32)$$

The corresponding standard free energy difference can be expressed as

$$\Delta G_p^0 = -k_{\mathrm{B}}T\ln K_p^{\mathrm{cage}} \approx p\Delta\mu_C^0, \quad (33)$$

$$\Delta\mu_C^0 \approx -k_B T\ln\left(\frac{v_t\omega_r}{8\pi^2}e^{\beta E_c}c_0\right), \quad (34)$$

for $p\gg 1$. Assuming that the simulated triskelia bound in a cage experience an estimated translational freedom of $0.1\sigma$ along every Cartesian direction and an estimated rotational freedom of 0.1 rad (~6°) around every Cartesian axis,

$$\Delta\mu_C^0 \approx 10.2\ k_{\mathrm{B}}T - E_c. \quad (35)$$

The resulting fraction of clathrin bound in cages, $f = p[C_p]/[C]_t$, is plotted in Fig. 11 as a function of the total clathrin concentration, $[C]_t = [C] + p[C_p]$. This fraction reaches a value of 50%, i.e., the number

of bound triskelia equals the number of unbound triskelia, when the total concentration equals the CAC; the first cages appear at approximately one-half this overall concentration. The above equilibrium constant can be related to the CAC, and thence to experimental data on clathrin. At the CAC, the number density of free triskelia reads as $[C] = c_{\mathrm{CAC}}/2$ and that of cages as $[C_p] = c_{\mathrm{CAC}}/(2p)$, hence

$$K_p^{\mathrm{cage}} = \frac{1}{p}\left(\frac{c_{\mathrm{CAC}}}{2c_0}\right)^{1-p}, \tag{36}$$

$$\Delta\mu_C^0 \approx k_{\mathrm{B}}T \ln\frac{c_{\mathrm{CAC}}}{2c_0} \tag{37}$$

for $p \gg 1$. The experimental CAC of 100 $\mu$g/mL (49) then translates into $\Delta\mu_C^0 \approx -16.4\ k_{\mathrm{B}}T$, and this value is reproduced by the simulation model for $E_c \approx 27\ k_{\mathrm{B}}T$. In simulations with an overall triskelion density close to the experimental CAC, the numbers of bound and unbound particles were approximately equal when using a leg-leg interaction strength $\epsilon \approx 6\ k_{\mathrm{B}}T$; the resulting average potential energy of ~23 $k_{\mathrm{B}}T$ per clathrin (33) is in good agreement with the above theoretical estimate. We note that the elementary volume elements $\Delta_C$ have again cancelled out in the statistical mechanical expression for the equilibrium constant. This was not the case in our earlier derivation, which consequently overestimated the binding energy (33). Muthukumar and Nossal (50) presented a derivation based on mole fractions, following the common practice in micelle theory (60), to arrive at an enthalpic energy $E_c \approx k_{\mathrm{B}}T \ln c_s/c_{\mathrm{CAC}} \approx 21\ k_{\mathrm{B}}T$, with the subscript $s$ referring to the solvent. There is no compelling physical reason to use mole fractions, and one now sees that the method works because the volume per solvent molecule, $v_s = 1/c_s$, provides a reasonable order of magnitude estimate for the libration volume of a clathrin bound in a cage, $v_t\omega_r$.

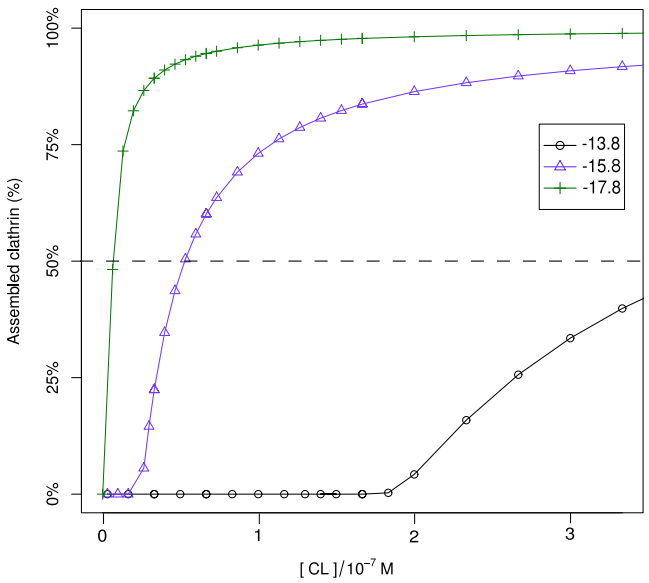


FIGURE 11 Theoretical concentration dependence of the fraction of particles bound in cages, for several values of the standard chemical potential difference $\Delta\mu_C^0$, indicated in units of $k_{\mathrm{B}}T$ in the legend, in the absence of APs. At the CAC, which varies with the interaction strength, the concentrations of free and bound clathrin are equal (*dashed line*). In this calculation, all cages are assumed of identical size, $p = 60$. Note the strong resemblance to the experimental data on in vitro assembly of clathrin cages (see Fig. 9 in Ungewickell and Ungewickell (62)). To see this figure in color, go online.

## APPENDIX III

## Decorated clathrin cages

Finally, we consider the formation of a cage decorated with $n$ single-clicked APs and $m$ double-clicked APs,

$$pC + (n+m)A \rightleftharpoons C_pA'_nA''_m. \tag{38}$$

To keep the derivation manageable, it is assumed that for every clicking site on a clathrin in a cage there is one nearest clicking site on an adjacent clathrin in that cage, such that the two sites—and hence the two triskelia—can be linked by an AP. Distance measurement reveal that the separation between two nearest sites on differing triskelia in a cage, $d_c$, is shorter than the distance $d_t$ between two nearest sites on the same triskelion. Because of the functional forms of $q_s''$ and $\psi$, a small reduction of the elongation of the entropic spring results in a pronounced increase of $q_s''$—we may therefore ignore intraclathrin double-clicked APs. Combining the results from the two preceding Appendices, we then arrive at the equilibrium constant

$$K_{p,n,m}^{\mathrm{cage}} = g_{p,n,m}\left(\frac{v_t\omega_r}{8\pi^2}\right)^{p-1}\left(\frac{4}{3}\pi\rho^3\right)^n\left(\frac{q_s''(d_c)}{q_s}\right)^m \times e^{\beta[pE_c+(n+2m)\zeta]}c_0^{p+n+m-1}, \tag{39}$$

where $n \geq 0$, $m \geq 0$, and $n + 2m \leq 6p$. Again, the elementary volume elements have cancelled out. The multiplicity is estimated as

$$g_{p,n,m} \approx g_p \frac{(3p)!}{m!(3p-m)!}\frac{(6p-2m)!}{n!(6p-2m-n)!}, \tag{40}$$

where the first factor, accounting for the cage structure, has been discussed before, $g_p \approx 1$; the second factor counts the permitted distributions of $m$ double-clicked APs over $3p$ pairs of nearest unlike click sites in a cage; and the third factor represents the permitted distributions of $n$ single-clicked APs over the remaining $6p - 2m$ free clicking sites of the cage. The standard free energy difference of the reaction can be expressed as

$$\Delta G_{p,n,m}^0 \approx p\Delta\mu_C^0 + n\Delta\mu_{A'}^0 + m\Delta\mu_{A''}^0(d_c) - k_{\mathrm{B}}T \ln g_{p,n,m}, \tag{41}$$

where the multiplicity is no longer negligibly small. The extension to site-dependent clicking strengths is again straightforward.

To obtain the number of cages at every point in the assembly diagrams of Figs. 7–10, we consider a closed system of volume $V$ with given total clathrin concentration $[C]_t$ and AP concentration $[A]_t$. For simplicity, we again consider only one cage size, $p = 60$. We denote the estimated concentrations of free, i.e., unbound, triskelia and APs as $[C]_f$ and $[A]_f$, respectively. The concentrations of decorated triskelia and decorated cages then follow by using the equilibrium constants derived above. A weighted sum over all species yields the sum concentrations of triskelia and APs present in the box,

$$[C]_s = [C]_f + \sum_{n,m} K_{n,m}^{\mathrm{tri}}[C]_f[A]_f^{n+m} + p\sum_{n,m} K_{p,n,m}^{\mathrm{cage}}[C]_f^p[A]_f^{n+m}, \tag{42}$$

$$[A]_s = [A]_f + \sum_{n,m}(n+m)K_{n,m}^{\mathrm{tri}}[C]_f[A]_f^{n+m} + \sum_{n,m}(n+m)K_{p,n,m}^{\mathrm{cage}}[C]_f^p[A]_f^{n+m}. \tag{43}$$

We now solve the equations $[C]_s = [C]_t$ and $[A]_s = [C]_t$ by varying $[C]_f$ and $[A]_f$. This is achieved by minimizing the sum-squared differences between the imposed and calculated overall concentrations, followed by a Newton-Raphson zero-point solver (61). The second summation on the right-hand side of Eq. 42 then yields the concentration of cages in the system. The presented calculations assume $d_c = 0.6\sigma$; because the limit underlying Eq. 20 does not hold true for the simulation model, the integral in Eq. 19 is solved numerically to yield $q_s''(d_c) \approx 7.8 \times 10^{-5}\sigma^6$. The standard chemical potential differences of single- and double-clicked APs are then related by

$$\Delta\mu^0_{A''}(d_c) \;=\; 2\Delta\mu^0_{A'} + 9.7\, k_B T, \tag{44}$$

while Eq. 27 predicts a slightly higher offset of 11.5 $k_BT$.

Measuring the hundreds of reaction equilibrium constants condensed into Eqs. 39 and 41 offers a daunting challenge, the more so as this requires identifying the numbers of triskelia, and single-clicked and double-clicked APs of every cage in the reaction mixture. To facilitate comparison of theory with experiment, we revert to a CAC. The condition where half the APs in a solution are bound to triskelia (both free triskelia and those in cages) is a complex function of all equilibrium constants of the mixture, as well as of the overall clathrin concentration, but it can be measured without detailed resolution of the compositions of aggregates. By measuring this CAC at a number of clathrin concentrations, it is in principle possible to extract all relevant standard chemical potential differences. An illustration hereof is provided in Fig. 12, where it is assumed that $\Delta\mu^0_C$ and the relation between $\Delta\mu^0_{A''}(d)$ and $\Delta\mu^0_{A'}$ are known in advance; by measuring AP's CAC at a given overall clathrin concentration, one can read the corresponding $\Delta\mu^0_{A'}$ from the graph.

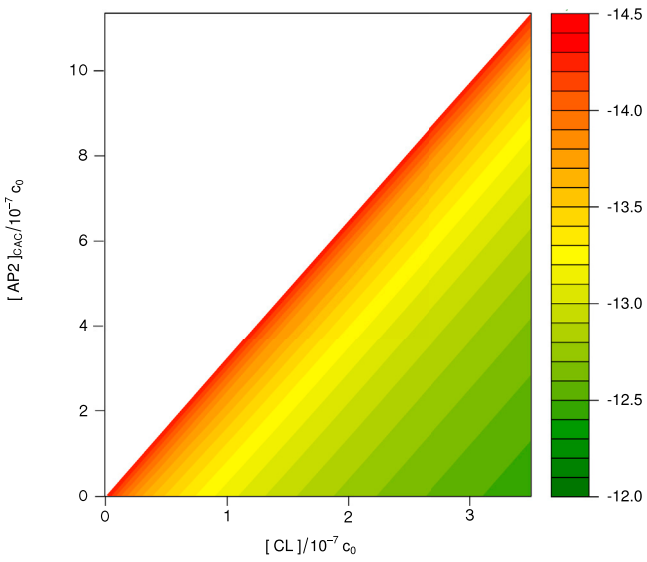


FIGURE 12 For any given total clathrin concentration (*horizontal axis*), there exists a total AP concentration (*vertical axis*) for which the fraction of bound AP equals the fraction of unbound AP. The plot relates both concentrations to the corresponding standard chemical potential difference for single-clicked APs, $\Delta\mu^0_{A'}$, in units of $k_BT$. The underlying calculation is based on $\Delta\mu^0_C = -11.8\, k_BT$ and assumes $\Delta\mu^0_{A''}$ and $\Delta\mu^0_{A'}$ to be related by Eqs. 28 and 44. The diagonal represents the maximum attainable AP CAC for a given clathrin concentration. To see this figure in color, go online.

## SUPPORTING MATERIAL

Supporting Materials and Methods, six figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30405-2.

## AUTHOR CONTRIBUTIONS

M.G., W.K.d.O., and W.J.B. designed the simulation model; M.G. implemented the simulation software, performed the simulations, and analyzed the data; W.K.d.O. developed the theoretical model; M.G. and W.K.d.O. wrote the article; and W.K.d.O. and W.J.B. supervised the study.

## ACKNOWLEDGMENTS

This work is part of the research program "Self-Assembly of Protein Coats at Membranes" (project No. 711.012.004), which is financed by the Netherlands Organisation for Scientific Research.

## REFERENCES

1. Royle, S. J. 2006. The cellular functions of clathrin. *Cell. Mol. Life Sci.* 63:1823–1832.
2. Young, A. 2007. Structural insights into the clathrin coat. . Seminars in Cell & Developmental Biology, Vol. 18. Elsevier, New York, pp. 448–458.
3. McMahon, H. T., and E. Boucrot. 2011. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12:517–533.
4. Brodsky, F. M. 2012. Diversity of clathrin function: new tricks for an old protein. *Annu. Rev. Cell Dev. Biol.* 28:309–336.
5. Kirchhausen, T., D. Owen, and S. C. Harrison. 2014. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb. Perspect. Biol.* 6:a016725.
6. Robinson, M. S. 2015. Forty years of clathrin-coated vesicles. *Traffic.* 16:1210–1238.
7. Brodsky, F. M., C.-Y. Chen, …, D. E. Wakeham. 2001. Biological basket weaving: formation and function of clathrin-coated vesicles. *Annu. Rev. Cell Dev. Biol.* 17:517–568.
8. Brodsky, F. M. 1988. Living with clathrin: its role in intracellular membrane traffic. *Science.* 242:1396–1402.
9. Kirchhausen, T. 1993. Coated pits and coated vesicles—sorting it all out. *Curr. Opin. Struct. Biol.* 3:182–188.
10. Ungewickell, E., and D. Branton. 1981. Assembly Units of Clathrin Coats. Nature Publishing Group, London, UK.
11. Pearse, B. M., C. J. Smith, and D. J. Owen. 2000. Clathrin coat construction in endocytosis. *Curr. Opin. Struct. Biol.* 10:220–228.
12. Maldonado-Báez, L., and B. Wendland. 2006. Endocytic adaptors: recruiters, coordinators and regulators. *Trends Cell Biol.* 16:505–513.
13. Schmid, E. M., and H. T. McMahon. 2007. Integrating molecular and network biology to decode endocytosis. *Nature.* 448:883–888.
14. Xing, Y., T. Böcking, …, S. C. Harrison. 2010. Structure of clathrin coat with bound Hsc70 and auxilin: mechanism of Hsc70-facilitated disassembly. *EMBO J.* 29:655–665.
15. Traub, L. M., and J. S. Bonifacino. 2013. Cargo recognition in clathrin-mediated endocytosis. *Cold Spring Harb. Perspect. Biol.* 5:a016790.
16. Keen, J. H. 1987. Clathrin assembly proteins: affinity purification and a model for coat assembly. *J. Cell Biol.* 105:1989–1998.
17. Kirchhausen, T. 2002. Clathrin adaptors really adapt. *Cell.* 109:413–416.
18. Owen, D. J., B. M. Collins, and P. R. Evans. 2004. Adaptors for clathrin coats: structure and function. *Annu. Rev. Cell Dev. Biol.* 20:153–191.

19. Robinson, M. S. 2004. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* 14:167–174.
20. Edeling, M. A., C. Smith, and D. Owen. 2006. Life of a clathrin coat: insights from clathrin and AP structures. *Nat. Rev. Mol. Cell Biol.* 7:32–44.
21. Popova, N., I. Deyev, and A. Petrenko. 2013. Clathrin-mediated endocytosis and adaptor proteins. *Acta Naturae.* 5:62.
22. Owen, D. J., Y. Vallis, …, P. R. Evans. 2000. The structure and function of the $\beta$ 2-adaptin appendage domain. *EMBO J.* 19:4216–4227.
23. Rappoport, J. Z. 2008. Focusing on clathrin-mediated endocytosis. *Biochem. J.* 412:415–423.
24. Collins, B. M., A. J. McCoy, …, D. J. Owen. 2002. Molecular architecture and functional model of the endocytic AP2 complex. *Cell.* 109:523–535.
25. Jackson, L. P., B. T. Kelly, …, D. J. Owen. 2010. A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell.* 141:1220–1229.
26. Kelly, B. T., S. C. Graham, …, D. J. Owen. 2014. Clathrin adaptors. AP2 controls clathrin polymerization with a membrane-activated switch. *Science.* 345:459–463.
27. Lindner, R. 1992. Interaktionen der Hüllstrukturproteine von Clathrin-Coated Vesicles. Technische Universität München, München, Germany.
28. Greene, B., S.-H. Liu, …, F. M. Brodsky. 2000. Complete reconstitution of clathrin basket formation with recombinant protein fragments: adaptor control of clathrin self-assembly. *Traffic.* 1:69–75.
29. Keen, J. H., K. A. Beck, …, T. Jarrett. 1991. Clathrin domains involved in recognition by assembly protein AP-2. *J. Biol. Chem.* 266:7950–7956.
30. Knuehl, C., C.-Y. Chen, …, F. M. Brodsky. 2006. Novel binding sites on clathrin and adaptors regulate distinct aspects of coat assembly. *Traffic.* 7:1688–1700.
31. den Otter, W. K., M. R. Renes, and W. J. Briels. 2010. Self-assembly of three-legged patchy particles into polyhedral cages. *J. Phys. Condens. Matt.* 22:104103.
32. den Otter, W. K., M. R. Renes, and W. J. Briels. 2010. Asymmetry as the key to clathrin cage assembly. *Biophys. J.* 99:1231–1238.
33. den Otter, W. K., and W. J. Briels. 2011. The generation of curved clathrin coats from flat plaques. *Traffic.* 12:1407–1416.
34. Ilie, I. M., W. K. den Otter, and W. J. Briels. 2014. Rotational Brownian dynamics simulations of clathrin cage formation. *J. Chem. Phys.* 141:065101.
35. Matthews, R., and C. N. Likos. 2013. Structures and pathways for clathrin self-assembly in the bulk and on membranes. *Soft Matter.* 9:5794–5806.
36. Cordella, N., T. J. Lampo, …, A. J. Spakowitz. 2014. Membrane fluctuations destabilize clathrin protein lattice order. *Biophys. J.* 106:1476–1488.
37. VanDersarl, J. J., S. Mehraeen, …, N. A. Melosh. 2014. Rheology and simulation of 2-dimensional clathrin protein network assembly. *Soft Matter.* 10:6219–6227.
38. Zaremba, S., and J. H. Keen. 1983. Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J. Cell Biol.* 97:1339–1347.
39. Fotin, A., Y. Cheng, …, T. Walz. 2004. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature.* 432:573–579.
40. Wilbur, J. D., P. K. Hwang, and F. M. Brodsky. 2005. New faces of the familiar clathrin lattice. *Traffic.* 6:346–350.
41. Böcking, T., F. Aguet, …, T. Kirchhausen. 2014. Key interactions for clathrin coat stability. *Structure.* 22:819–829.
42. Tompa, P. 2009. Structure and Function of Intrinsically Disordered Proteins. Chapman and Hall/CRC, Boca Raton, FL.
43. Boal, D. 2002. Mechanics of the Cell. Cambridge University Press, Cambridge, UK.
44. Kremer, K., and G. S. Grest. 1990. Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. *J. Chem. Phys.* 92:5057–5086.
45. Frenkel, D., and B. Smit. 2001. Understanding Molecular Simulation, 2nd Ed. Academic Press, Orlando, FL.
46. Landau, D. P., and K. Binder. 2014. A Guide to Monte Carlo Simulations in Statistical Physics. Cambridge University Press, Cambridge, UK.
47. Allen, M. P., and D. J. Tildesley. 1989. Computer Simulation of Liquids. Clarendon Press, New York.
48. Swendsen, R. H., and J.-S. Wang. 1987. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* 58:86–88.
49. Crowther, R. A., and B. M. Pearse. 1981. Assembly and packing of clathrin into coats. *J. Cell Biol.* 91:790–797.
50. Muthukumar, M., and R. Nossal. 2013. Micellization model for the polymerization of clathrin baskets. *J. Chem. Phys.* 139:121928.
51. Saleem, M., S. Morlot, …, A. Roux. 2015. A balance between membrane elasticity and polymerization energy sets the shape of spherical clathrin coats. *Nature* Comm. 6:6249.
52. Schein, S., and M. Sands-Kidner. 2008. A geometric principle may guide self-assembly of fullerene cages from clathrin triskelia and from carbon atoms. *Biophys. J.* 94:958–976.
53. Schoen, A. P., N. Cordella, …, S. C. Heilshorn. 2013. Dynamic remodelling of disordered protein aggregates is an alternative pathway to achieve robust self-assembly of nanostructures. *Soft Matter.* 9:9137–9145.
54. Nelson, P. 2004. Biological Physics. WH Freeman, New York.
55. McQuarrie, D. A. 1973. Statistical Thermodynamics. HarperCollins Publishers, New York.
56. Phillips, R., J. Kondev, …, H. Garcia. 2012. Physical Biology of the Cell. Garland Science Publishing, Hamden, CT.
57. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.
58. Ben-Tal, N., B. Honig, …, A. Ben-Shaul. 2000. Association entropy in adsorption processes. *Biophys. J.* 79:1180–1187.
59. Schein, S., M. Sands-Kidner, and T. Friedrich. 2008. The physical basis for the head-to-tail rule that excludes most fullerene cages from self-assembly. *Biophys. J.* 94:938–957.
60. W. M. Gelbart, A. Ben-Shaul, and D. Roux, editors 1994. Micelles, Membranes, Microemulsions and Monolayers. Springer, Berlin, Germany.
61. Press, W. H., S. Teukolsky, …, B. Flannery. 1992. Numerical Recipes in FORTRAN. Cambridge University Press, Cambridge, UK.
62. Ungewickell, E., and H. Ungewickell. 1991. Bovine brain clathrin light chains impede heavy chain assembly in vitro. *J. Biol. Chem.* 266:12710–12714.

**Biophysical Journal, Volume *111***

## Supplemental Information

# Clathrin Assembly Regulated by Adaptor Proteins in Coarse-Grained Models

**Matteo Giani, Wouter K. den Otter, and Wim J. Briels**

# Supplementary Material

# Clathrin assembly regulated by adaptor proteins in coarse-grained models

M. Giani, [1,2,3] W.K. den Otter [1,2,3] and W.J. Briels [2,3,4]

[1] Multi Scale Mechanics, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[2] Computational BioPhysics, Faculty of Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[3] MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[4] Forschungszentrum Jülich, ICS 3, D-52425 Jülich, Germany

## The clathrin model

This supplementary material describes the main features of our coarse-grained clathrin model. We refer the interested reader to earlier publications for more details and a motivation of the model.

### Shape

The particle is rigid. Each leg consists of a three linear segments, referred to as the proximal ($p$), distal ($d$) and terminal domain ($td$), see Fig 1. The three proximal segments meet at the central hub ($h$), at a pucker angle $\chi = 101°$ relative to the normal vector $\hat{\mathbf{n}}_h$ along to the three fold symmetry axis of the particle. Normal vectors are also associated with every knee ($k$); the normal $\hat{\mathbf{n}}_{\alpha,k}$ of the knee of the $\alpha^{\text{th}}$ leg lies in the plane formed by the proximal segment of that leg and the particle normal $\hat{\mathbf{n}}_h$. The three angles between a knee-normal and the two leg segments meeting at that knee are chosen to be identical to the three angles between the hub-normal and two proximal segments meeting at the hub. The terminal domain is attached at the ankle ($a$) at an angle of 114° relative to the adjacent distal segment, and at a dihedral angle of 28° relative to the distal and proximal segments of the same leg. All leg segment are taken to be the same length, $\sigma = 17\,\text{nm}$. Associated with every proximal segment is a polarity vector, defined for the $\alpha^{\text{th}}$ leg as

$$\hat{\mathbf{m}}_{\alpha,p} = \frac{(\mathbf{x}_{\alpha,k} - \mathbf{x}_h) \times \hat{\mathbf{n}}_h}{|(\mathbf{x}_{\alpha,k} - \mathbf{x}_h) \times \hat{\mathbf{n}}_h|}, \tag{1}$$

with $\mathbf{x}$ denoting the position of the specific joint indicated in the subscript, see Fig. S1. Polarity vectors to distal segment $\hat{\mathbf{m}}_{\alpha,d}$ are defined likewise, based on the end points of that segment and the normal at the knee.

### Potential

The interaction between two triskelia is described by a sum of inter-segmental interactions. These interactions conform with the segmental pairings observed in experimental cage

edges: attractive interactions are introduced between aligned pairs of two anti-parallel proximal segments, of two anti-parallel distal segments, and between any aligned pair of one proximal and one distal segment; here 'anti-parallel' refers to amino acid sequences running in opposite directions. When two segments are properly aligned, their respective ends are close to each other. As an example, consider the proximal segment of the $\alpha^{\text{th}}$ leg of particle $i$ interacting with the proximal segment of the $\beta^{\text{th}}$ leg of particle $j$. The two average distances between the four ends of these segments are

$$r_{j\beta,kh}^{i\alpha,hk} = \frac{1}{2}\left|\mathbf{x}_{i,h} - \mathbf{x}_{j\beta,k}\right| + \frac{1}{2}\left|\mathbf{x}_{i\alpha,k} - \mathbf{x}_{j,h}\right|, \tag{2}$$

$$r_{j\beta,hk}^{i\alpha,hk} = \frac{1}{2}\left|\mathbf{x}_{i,h} - \mathbf{x}_{j,h}\right| + \frac{1}{2}\left|\mathbf{x}_{i\alpha,k} - \mathbf{x}_{j\beta,k}\right|, \tag{3}$$

where the distance on the first line is small if the hub of $i$ is close to the $\beta^{\text{th}}$ knee of $j$ and the $\alpha^{\text{th}}$ knee of $i$ is close to the hub of $j$ (*i.e.* aligned and anti-parallel), while the distance on the second line is small if the hub of $i$ is close to the hub of $j$ and the $\alpha^{\text{th}}$ knee of $i$ is close to the $\beta^{\text{th}}$ knee of $j$ (*i.e.* aligned and parallel), as illustrated in Fig. S2. The former combination occurs in clathrin cages, hence an attractive interaction is assigned:

$$\phi_{j\beta,kh}^{i\alpha,hk} = -\varepsilon_{kh}^{hk} \cdot f\left(r_{j\beta,kh}^{i\alpha,hk}\right) \cdot g\left(\hat{\mathbf{m}}_{i\alpha,p} \cdot \hat{\mathbf{m}}_{j\beta,p}\right), \tag{4}$$

with the positive parameter $\varepsilon_{kh}^{hk}$ denoting the (absolute) maximum inter-segmental binding energy. The distance dependence smoothly decreases from unity for coinciding end points to zero at the cut-off distance $r_{\text{cut}}$, following

$$f(r) = \frac{1}{2}\left[1 - \frac{\tanh[A(r - r_{\text{cut}}/2)]}{\tanh[Ar_{\text{cut}}/2]}\right], \tag{5}$$

where $A$ determines the steepness of the potential, see Fig. S3. The numerical values entering these expressions are provided in Table S1. In addition to a small distance between the end points, we also require alignment of the polarities of the two leg segments, through

$$g(x) = \begin{cases} -x & \text{for} \quad x < 0 \\ 0 & \text{for} \quad x \geq 0. \end{cases} \tag{6}$$

This factor reflects the supposition that interaction sites are not distributed homogeneously over the segmental surface, but are concentrated on the side that faces the three neighbouring segments in a cage edge. As a result, two legs will bind only if the two triskelia involved are oriented similarly, *i.e.* their hub normals point approximately in the same direction. For the current example of two proximal segments, they will bind only if the average hub-knee distance is small and their polarities are pointing in opposite directions. Our earlier simulations indicated that this rotational asymmetry holds the key to the self-assembly of clathrin cages; upon removing the factor $g$ from the potential, triskelia form disordered aggregates rather than cages. All other attractive interactions between binding segment pairs, see the aforementioned list of combinations, are constructed along the same lines.

The introduction of excluded volume interactions between the thin long legs is disadvantageous from a computational point of view: they impose a smaller time step, and the legs have to be non-linear and slightly flexible in order to interweave into a cage edge. We therefore omitted excluded volume interactions. One important consequence of excluded volume interactions should not be ignored, however: excluded volume prevents a leg segment from binding to an edge in a 'slot' already occupied by another leg segment.

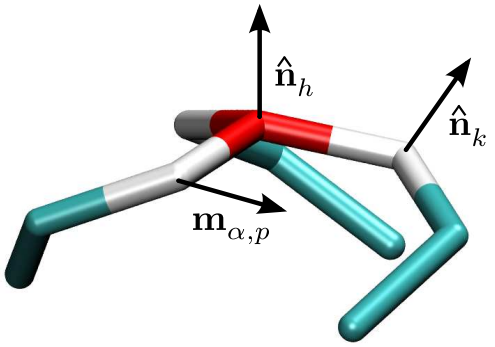


Figure S1: The coarse grained simulation model. The unit vectors $\hat{\mathbf{n}}_h$ and $\hat{\mathbf{m}}_{\alpha,p}$ denote the normal at the hub and the polarity vector of a proximal segment, respectively.

Returning to the two proximal segments of the above example: the distance between the ends of two aligned parallel proximal segments, as calculated in Eq. (3), should not become small. This is achieved by a pair interaction similar to Eq. (4),

$$\phi^{i\alpha,hk}_{j\beta,kh} = -\varepsilon^{hk}_{hk} \cdot f\left(r^{i\alpha,hk}_{j\beta,hk}\right), \tag{7}$$

where the interaction strength parameter $\varepsilon^{hk}_{hk}$ has a negative value and the rotational asymmetry has been omitted. A likewise interaction is introduced between two aligned parallel distal segments. The parameters of these repulsive interactions are listed in Table S1.

| $i\alpha - j\beta$ | $\varepsilon$ | $A/\sigma^{-1}$ | $r_{\text{cut}}/\sigma$ | $g(x)$ |
|---|---|---|---|---|
| **attractive** | | | | |
| $hk - kh$ | $\epsilon$ | 4 | 0.4 | $\hat{\mathbf{m}}_p \cdot \hat{\mathbf{m}}_p$ |
| $ka - ak$ | $\epsilon$ | 4 | 0.4 | $\hat{\mathbf{m}}_d \cdot \hat{\mathbf{m}}_d$ |
| $hk - ka$ | $\epsilon/2$ | 4 | 0.4 | $-\hat{\mathbf{m}}_p \cdot \hat{\mathbf{m}}_d$ |
| $hk - ak$ | $\epsilon/2$ | 4 | 0.4 | $\hat{\mathbf{m}}_p \cdot \hat{\mathbf{m}}_d$ |
| **repulsive** | | | | |
| $hk - hk$ | $-10\epsilon$ | 0.8 | 0.8 | $-1$ |
| $ka - ka$ | $-10\epsilon$ | 0.8 | 0.8 | $-1$ |

Table S1: Interaction parameters of the six distinct clathrin leg segment pairings. In the first column, the letters refer to the hub ($h$), knee ($k$) and ankle ($a$) of legs $\alpha$ and $\beta$ of particles $i$ and $j$, respectively. Note that the order is important: the two proximal-proximal pairings, *i.e.* the first attractive combination and the first repulsive combination, refer to Eqs. (2) and (3), respectively. The value of $\epsilon$ is varied from 2 to $10k_BT$ in the construction of the phase diagrams. The elements in the last column represent the arguments $x$ to the polarity function $g(x)$, where the first vector in the dot products refers to a segment of the $\alpha$ leg of particle $i$ and the second vector to a segment of the $\beta$ leg of particle $j$, and where $g(-1) = 1$.

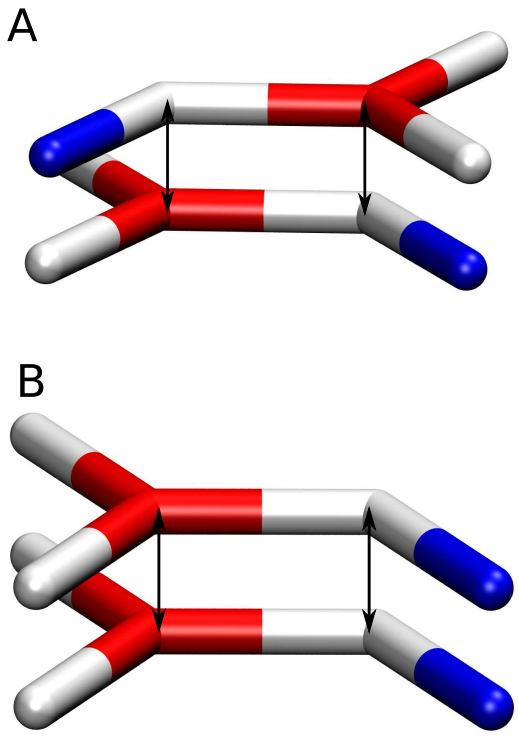


Figure S2: (A) Arrows indicating the two distances entering Eq. (2), to be used for the attractive potential, and (B) arrows indicating the two distances entering Eq. (3), to be used for the repulsive potential.

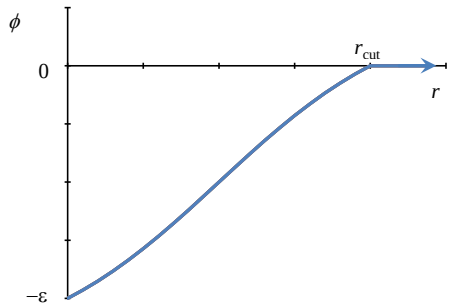


Figure S3: Distance dependence of the segment-segment interaction.

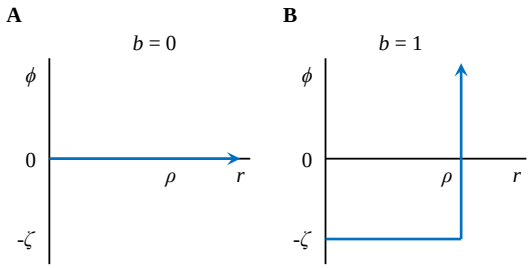


Figure S4: AP-clathrin click potential in the unclicked (A, $b = 0$) and clicked (B, $b = 1$) state.

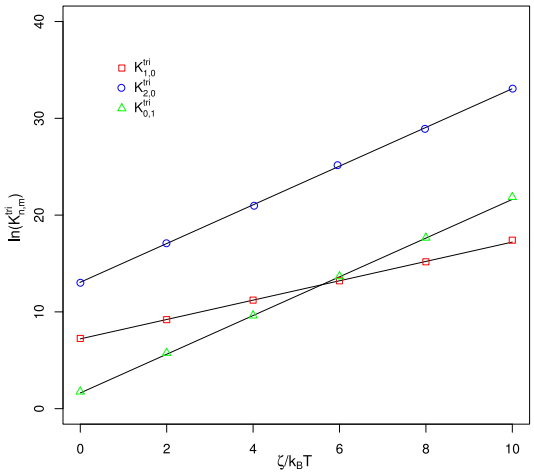


Figure S5: Validation of the Monte Carlo clicking moves by comparing reaction equilibrium constants obtained from simulations (markers) with the theoretical expressions (solid lines) derived in Appendix 1. Shown are the reaction equilibrium constants for triskelia with one single-clicked AP, $K^{\text{tri}}_{1,0}$, with two single-clicked APs, $K^{\text{tri}}_{2,0}$, and with one double-clicked AP, $K^{\text{tri}}_{0,1}$. Simulations are performed using a box of volume $10^6\sigma^3$ populated with 1,000 clathrin and 3,000 APs. To enhance the number of clicks and to facilitate the comparison with theory, the clathrin-clathrin interactions are turned off, $\epsilon = 0$, AP clicks are limited to one leg per triskelion, the AP spring constant is reduced to $k = 1k_BT/\sigma^2$, and the radius of the click interaction is enlarged to $\rho = 0.3\sigma$.

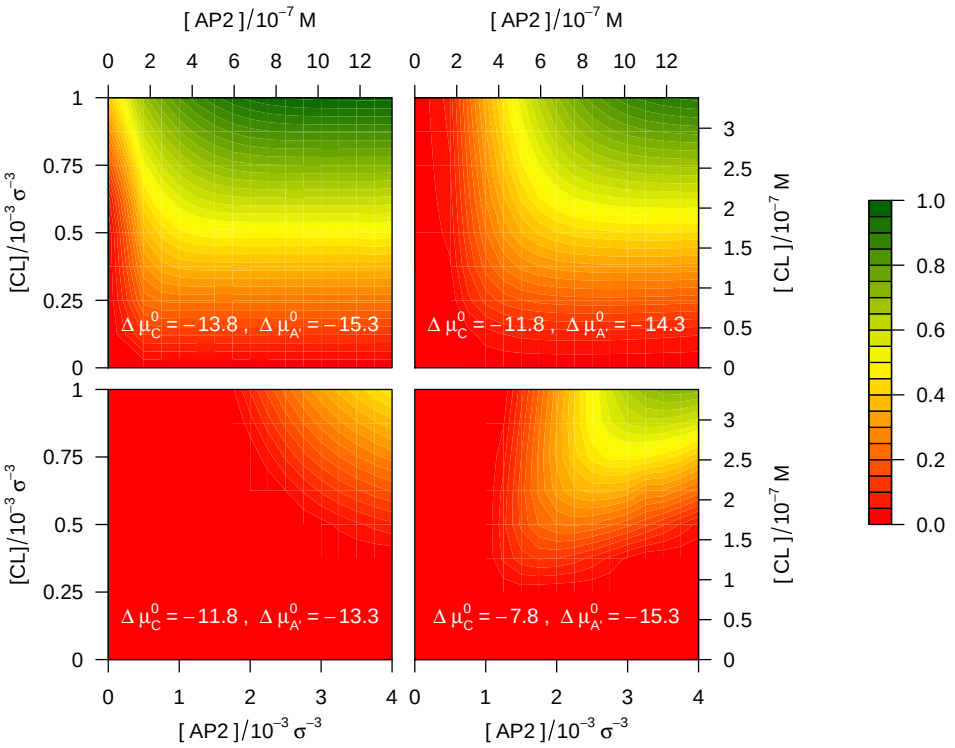


Figure S6: Calculated fraction of triskelia absorbed in cages, for APs clicking to the ends of the TDs and the ankles of triskelia, as a function of the clathrin and AP concentrations, for the standard chemical potential differences indicated on the plots, in units of $k_BT$. See main text for discussion.