Self-assembly of three-legged patchy particles into polyhedral cages

2010 J. Phys.: Condens. Matter 22 104103

(http://iopscience.iop.org/0953-8984/22/10/104103)

J. Phys.: Condens. Matter **22** (2010) 104103 (8pp) doi:10.1088/0953-8984/22/10/104103

Self-assembly of three-legged patchy particles into polyhedral cages

Wouter K den Otter, Marten R Renes and W J Briels

Computational Biophysics Group, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

E-mail: w.k.denotter@utwente.nl and w.j.briels@utwente.nl

Received 12 October 2009, in final form 12 January 2010
Published 23 February 2010
Online at stacks.iop.org/JPhysCM/22/104103

Abstract

The self-assembly of rigid three-legged building blocks into polyhedral cages is investigated by patchy particle simulations. A four-site anisotropic interaction potential is introduced to make pairs of overlapping legs bind in an anti-parallel fashion, thereby forming the edges of a polyhedron of pentagons and hexagons. A torsional potential, reflecting an asymmetry or polarity in the legs' binding potential, proves crucial for the successful formation of closed fullerene-like cages. Self-assembly proceeds by a nucleation-and-growth mechanism, with a high success rate of cage closure. The size distribution of the self-assembled buckyballs is largely determined by the pucker angle of the particle. Nature explores a similar building block, the clathrin triskelion, to regulate vesicle formation at the cell membrane during endocytosis.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The last decade has witnessed the development of 'patchy particles', large molecules or nanocolloids decorated with specific interaction sites to target their self-assembly into aggregates with desirable architectures [1]. These advanced particles offer the potential to grow materials with specific structures and according special properties, like diamond lattices of colloidal particles to create photonic bandgaps [2]. Another promising application is the development of nano-scale cargo containers that release their contents after a controlled trigger, e.g. a change in temperature or pH, or upon entering a diseased cell. Experiments with cleverly designed DNA strings have already been used for the 'bottom-up fabrication' of complex two and three dimensional structures [3, 4].

The self-assembly properties of patchy particles are heavily relied upon as a building principle in Nature. A well-known example is the lipids, whose combination of a hydrophilic head group and a hydrophobic tail makes them spontaneously aggregate into membranes that envelope cells and cell organelles. Another interesting example is the protein capsids by which viruses transport their genes from one cell to the next. These examples have in common that their assembly is readily reproduced and reversible *in vitro*, while the sizes of their aggregates differ widely—capsid proteins often form monodisperse cages of icosahedral symmetry [5, 6] whereas membranes can adopt virtually any size [7, 8]. An intermediate example is clathrin, a protein that uses its three long legs to self-assemble scaffolds next to a lipid membrane, thereby curving the membrane to eventually create an encaged vesicle [9–12]. Clathrin plays an important structural and regulatory role in the production of cargo-laden vesicles at the cell membrane (during endocytosis) and at the trans-Golgi network (during exocytosis). Depending on the size of the cargo, twenty up to several hundreds of these curved and slightly flexible triskelia self-assemble into wire-cages with twelve pentagonal and a variable number of hexagonal faces, with the centre of a triskelion positioned at every vertex and with edges consisting of four leg segments bound together by multiple weak interactions [13, 14]. The question of how the triskelia manage to self-assemble these complex cages, which proceeds *in vitro* under slightly acidic conditions without being facilitated by other proteins or membranes, continues to intrigue biologists and biophysicists. While clathrin cages are similar to buckyballs in connectivity [15], both having vertices bonded to three neighbouring vertices to form a lattice of pentagons and hexagons, it is evident that clathrins exploit markedly different interactions.

Computer simulations are increasingly used to investigate self-assembly processes of patchy particles. This requires endowing the simulated building blocks with anisotropic or

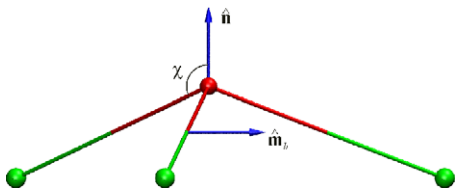

Figure 1. Cartoon of the rigid three-legged patchy particle, with the central A site (red) and three B sites (green) marked by spheres. The particle's normal $\hat{\mathbf{n}}$ points along the symmetry axis, and in a cage will be directed from the centre of the cage outwards. The leg's polarity vectors $\hat{\mathbf{m}}_b$, of which only one is drawn for clarity, are constructed following equation (5).

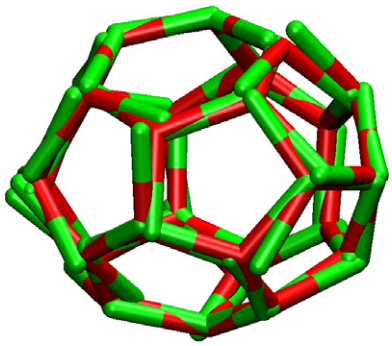

Figure 2. Snapshot of the smallest self-assembling cage, a dodecahedron containing twenty particles, illustrating the anti-parallel alignment of the two legs along every edge. The legs are coloured in red and green to highlight the positions of the A and B sites, respectively. The legs of a clathrin triskelion are curved and twice as long, enabling them to reach from one vertex to the next-nearest vertex and consequently results in four legs running along every edge.

directed interaction potentials, which is achieved through two alternative routes. In some studies the anisotropic building blocks are constructed as rigid collections of interaction sites with isotropic Lennard-Jones interactions [16–21]. This approach has the advantages that conventional simulation codes can still be used and it is fairly straightforward to make complex shapes, such as spheres with multiple patches, cones and triangular plates. Anisotropic interactions have also been obtained by multiplying a Lennard-Jones potential with a function of the relative orientations and relative positions of the particles [22–28], which has the advantage of strongly reducing the number of interaction sites. Computer simulations with these models contribute to understanding the basic building principles behind the self-assembly of simple and complex particles, as well as providing an ideal test ground for the analysis and development of new particle structures without the cumbersome process of synthesizing the new particle.

In this study, we simulate the self-assembly of rigid three-legged patchy particles, as illustrated in figure 1, into various large polyhedral cages. The model is inspired by the clathrin triskelion and by the notion that patchy particles offer the first promise of simulating the self-assembly of clathrin cages. For simplicity, we here assume that the legs are straight and hence there will be only two anti-parallel legs running along every lattice edge, see figure 2. A minimum model with the desired building capacities is introduced in sections 2 and 3, where it is shown that a remarkably simple force field suffices. The dependence of the assembly process, and the distribution of cages, on the details of the simulation model are discussed in section 4. The main conclusions of this study are summarized in section 5, along with an elaboration on their implications for the clathrin protein and man-made triskelia with similar assembly properties.

2. Model

With the complex clathrin triskelion and its intriguing self-assembly capabilities in mind, we here take a reductionist's point of view and construct a patchy particle model with three straight thin legs that bind in anti-parallel alignment. The simplest building block design, in our opinion, then involves a combination of both patchy particle approaches mentioned in section 1, namely a rigid unit with several anisotropic interaction sites. Our simulation model, illustrated in figure 1, comprises three straight thin legs of length σ, each starting at the central A site with the same 'pucker' angle χ relative to the symmetry axis $\hat{\mathbf{n}}$ and terminating at a B site. Since the small excluded volumes of the thin legs are expected to be of limited consequence to the thermodynamic properties, are a considerable complication to the assembly process, are time consuming to calculate numerically and are laborious to implement in a simulation code [29], we have chosen not to include them in the current model but to keep the particles as simple as possible for the time being.

The anti-parallel binding of the two legs forming a lattice edge, see figure 2, places the bth B site of particle i near the A site of particle j and simultaneously places the b'th B site of particle j near the A site of particle i, with $b, b' \in \{1, 2, 3\}$. The resulting large contact area between the two legs results, for clathrin triskelia, in many weak interactions that collectively from a strong leg–leg interaction, with the distribution of the interaction sites and/or their discriminating binding properties giving rise to the preference for anti-parallel binding. One may, of course, reproduce these binding characteristics in the simulations by modelling the legs as series of weak interaction sites, but a similar result is achieved at considerably lower computational costs by the introduction of a four-site interaction potential:

$$\phi = -\epsilon f(\tfrac{1}{2}|\mathbf{x}_{ib} - \mathbf{x}_{j0}| + \tfrac{1}{2}|\mathbf{x}_{jb'} - \mathbf{x}_{i0}|). \tag{1}$$

Here ϵ is a positive number determining the strength of the bond, $\mathbf{x}_{i0}$ and $\mathbf{x}_{ib}$ are the positions of the A and bth B site of particle i, respectively, and the vertical bars $|\cdots|$ denote the length of a vector. The dependence of the potential on the mean distance $r_{ib,jb'}$ between the unlike sites of the bth leg of particle i and the b'th leg of particle j is conveniently chosen as

$$f(r_{ib,jb'}) = \frac{\tanh[-\alpha(r_{ib,jb'} - r_c/2)]}{2\tanh[\alpha r_c/2]} + \frac{1}{2} \qquad \text{for } r_{ib,jb'} \leqslant r_c \tag{2}$$

and $f(r_{ib,jb'}) = 0$ beyond the cut-off distance r_c. As illustrated in figure 3, the function f smoothly passes from unity at

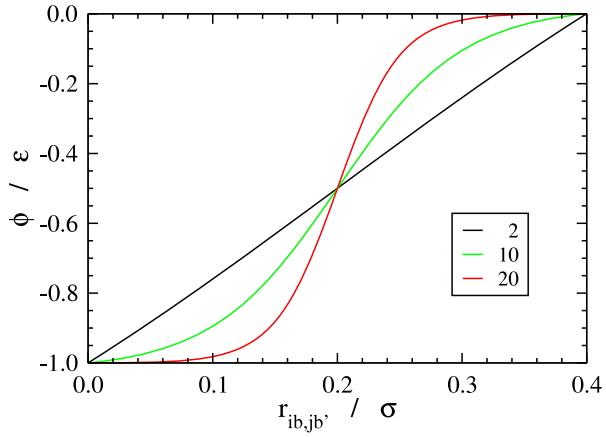

Figure 3. Plot of the four-site interaction potential ϕ, as defined in equations (1) and (2), against the average distance $r_{ib,jb'}$ between the bth leg of particle i and the b'th leg of particle j, for a cut-off radius $r_c = 0.4\sigma$, at three values of α (in units of σ^{-1}).

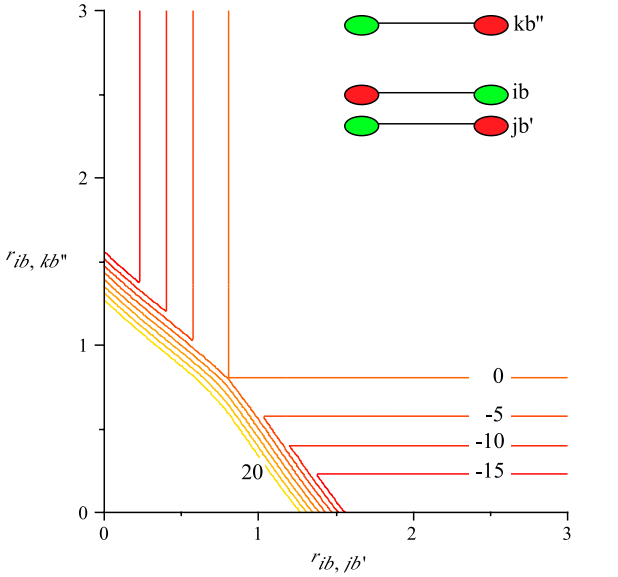

Figure 4. The potential energy surface for a set of three co-planar aligned legs, see inset, as a function of the distances $r_{ib,jb'}$ and $r_{ib,kb''}$ separating the anti-parallel pairs. The parameters of the repulsion between the parallel legs jb' and kb'' have been tuned to permit the two anti-parallel legs ib and jb' to bind when the second parallel leg kb'' is remote, while preventing binding of ib with either anti-parallel leg if jb' and kb'' approach each other too closely. A cross-section of the potential along $r_{ib,jb'}$, at constant $r_{ib,kb''} \geqslant 1.6\sigma$, yields the attractive potential of figure 3 for the interacting $ib - jb'$ pair; a similar result is obtained for the $ib - kb''$ pair by a cross-section at constant $r_{ib,jb'} \geqslant 1.6\sigma$.

zero distance, through a symmetry point at $r_c/2$, to zero at the cut-off distance, with the parameter α tuning the shape from a straight line for small α to a step function for large α. This potential combines the desired properties—pairs of legs bind with maximum overlap of their contact areas and in anti-parallel orientation—by bringing together the opposite ends of the legs. Note that changes in the angle between two neighbouring legs, or a sliding motion of one leg along the other, will affect the binding energy.

In the absence of excluded volume interactions, two anti-parallel legs binding together do not exclusively occupy and thereby monopolize their counterpart's binding capabilities. Hence, a third aligned leg (the b''th leg of particle k in the sketch of figure 4) approaching an anti-parallel pair of legs will bind to the oppositely aligned leg (leg b of particle i) of the pair, unopposed by the co-aligned leg (leg b' of particle j) of the pair. This undesired form of several legs binding along a lattice edge occurs readily in simulations with the purely attractive potential ϕ. To suppress triplet formation, and thereby multiplet formation, we want to create a situation whereby the repulsion between the third leg (kb'') and its parallel counterpart of the pair (jb') outweighs the attraction between the third leg (kb'') and the anti-parallel leg of the pair (ib). This is achieved by supplementing the attractive interactions ϕ with a repulsive four-site potential $\bar{\phi}$, which we also choose to be of the form of equations (1) and (2), but now based on the average distance between the like interaction sites of the two competing legs,

$$\bar{r}_{jb',kb''} = \tfrac{1}{2}|\mathbf{x}_{j0} - \mathbf{x}_{k0}| + \tfrac{1}{2}|\mathbf{x}_{jb'} - \mathbf{x}_{kb''}|. \qquad (3)$$

Here, bars are introduced to distinguish the argument $\bar{r}$ and parameters $\bar{\epsilon}$, $\bar{\alpha}$ and $\bar{r}_c$ of the repulsive potential $\bar{\phi}$ from their counterparts in the attractive potential ϕ, with both potentials having the same functional form. From the potential energy surface for three legs, as illustrated in figure 4, we conclude that suitable conditions are obtained for $\bar{\epsilon} = -10\epsilon$, $\bar{\alpha} = \alpha/5$ and $\bar{r}_c = 2r_c$. The total potential energy of a collection of N particles is then calculated as

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{b=1}^{3} \sum_{b'=1}^{3} \{\phi(r_{ib,jb'}) + \bar{\phi}(\bar{r}_{ib,jb'})\}, \qquad (4)$$

where the summation runs over all leg–leg combinations and includes both the parallel and anti-parallel orientation for every pair of legs. Note that, for the employed simulation parameters, the attractive and repulsive potentials are mutually exclusive for any given pair of legs.

The self-assembly process is simulated using the Metropolis Monte Carlo method [30, 31]. In every Monte Carlo (MC) step, the centre of mass of a randomly selected particle is subjected to a random 3D translation to a new position within a box of sides 0.25σ, centred around its old position, followed by a rotation around a random axis by a random angle between plus and minus 0.5 rad. The new configuration is then accepted with a probability $p = \min[\exp(-\beta\Delta\Phi), 1]$, where $\Delta\Phi$ denotes the energy change incurred in the trial move and $\beta = (k_B T)^{-1}$ with Boltzmann's constant k_B and temperature T. A grid list [30, 31] and a look-up list of leg–leg energies are used to accelerate the calculation of $\Delta\Phi$, with updates added after every accepted trial move. A typical simulation follows $N = 1000$ or $N = 10\,000$ particles at a number density $\rho = 10^{-3}\sigma^{-3}$ in a cubic box with periodic boundary conditions. The interaction cut-off r_c is set at 0.4σ and the steepness coefficient α to $2\sigma^{-1}$. The leg–leg attraction strength ϵ has been varied from $2k_B T$ to $50k_B T$, and the pucker

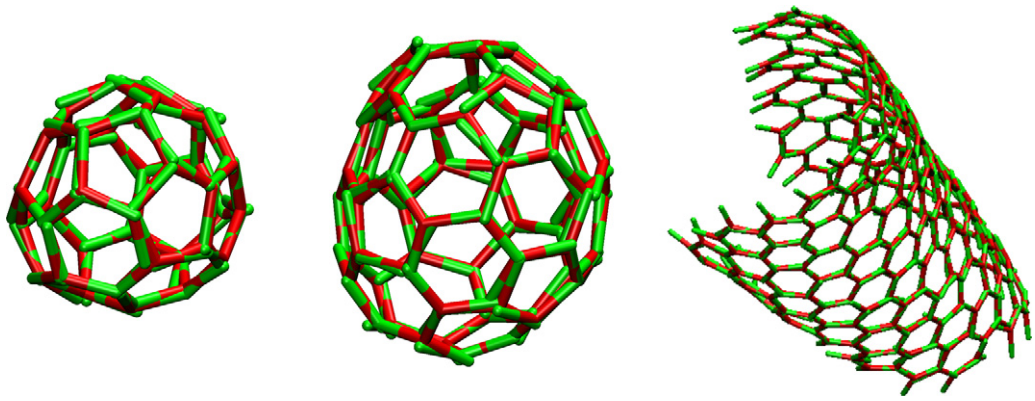

Figure 5. Snapshots showing several self-assembled structures. The 'hexagonal barrel' of 36 particles (left) is produced by triskelia with a pucker angle of 104°, the buckyball of 60 particles (middle) evolves from triskelia with $\chi = 101°$. The conical fragment of 'chicken-wire' contains particles with $\chi = 95°$. The strong correlation between pucker angle and cage size, see also figures 7 through 9, is discussed in the main text.

angle χ from 90° to 110°, to explore their influences on the assembly process.

3. Polarity

The simulations show that the particles readily aggregate into clusters of various sizes, provided ϵ exceeds a threshold binding energy, but the assembled structures are highly irregular in shape. Varying the parameters of the model did not resolve this issue, thereby naturally leading to the conclusion that self-assembly into closed cages requires an extension of the force field. Visual inspection of the grown structures, using the visual molecular dynamics (VMD) package [32], confirmed that the legs bind in an anti-parallel fashion, with exactly two legs per lattice edge, but that the torsional angles around these edges vary wildly. The symmetry axes or normals $\hat{\mathbf{n}}_i$ of neighbouring particles show little correlation in the sprawling aggregates, whereas they ought to be nearly parallel in order to form a closed cage. We therefore now augment the model with an anisotropy in the leg–leg interaction, to promote the orientational alignment of neighbouring particles.

In the extended model, each leg is endowed with a *polarity* vector $\hat{\mathbf{m}}_{ib}$ to introduce a rotational asymmetry in the leg–leg interactions. This solution is motivated by the experimental observations that clathrins in a cage consistently present the same faces of their legs to their neighbours, and hence that their interaction sites are non-uniformly distributed over the surface of the legs [9–14]. We here make the arbitrary choice of assigning front–back asymmetry to the legs, see figure 1, by defining the polarity vector of the bth leg of particle i as the cross product of the particle's normal with the leg's end-to-end vector,

$$\hat{\mathbf{m}}_{ib} = \frac{\hat{\mathbf{n}}_i \times (\mathbf{x}_{ib} - \mathbf{x}_{i0})}{|\hat{\mathbf{n}}_i \times (\mathbf{x}_{ib} - \mathbf{x}_{i0})|}. \tag{5}$$

In an ordered cage, as depicted in figure 2, the two anti-parallel legs in an edge have their polarities pointing in opposite directions. This suggests modifying the attractive potential of equation (1) into a direction specific interaction of the form

$$\phi = -\epsilon f(r_{ib,jb'}) g(\hat{\mathbf{m}}_{ib} \cdot \hat{\mathbf{m}}_{jb'}), \tag{6}$$

where the variation with the dot product μ of the polarities is chosen as

$$g(\mu) = \begin{cases} 0 & \text{for } \mu > 0 \\ -\mu & \text{for } \mu \leqslant 0. \end{cases} \tag{7}$$

This potential induces a preference for anti-parallel legs to bind with their polarities in anti-parallel fashion, i.e. with the largest negative μ, and thereby strongly promotes successful cage assembly over random aggregation, as will be discussed in the next section, while anti-parallel legs with parallel polarities, i.e. positive μ, do not bind at all. Dihedral interactions are also being employed in patchy particle simulations of virus capsid assembly, either implicitly by modelling a capsid protein as a rigid unit of several particles [18, 20, 21] or explicitly as part of the interaction potential between particles representing one capsid protein each [22, 24], but the crucial importance of the dihedral for the assembly process has thus far received little attention. Note that the repulsive potential between parallel legs remains independent of the legs' polarities.

4. Results

The model with anisotropic leg–leg interactions proves very capable of self-assembly into closed polyhedral cages. Several typical cages, grown at different pucker angles, are shown in figures 2 and 5. The closed cages have F_5 pentagonal and F_6 hexagonal faces; although other faces are in principle also possible, the only exceptions detected by our analysis software throughout all simulations are a few heptagons. One readily shows that the number of vertices equals the number of particles, $V = N$, that the number of edges is given by $E = 3N/2$, and that this number is related to the number of faces by $E = (5F_5 + 6F_6)/2$. In combination with Euler's famous polyhedron formula, $V - E + F = 2$, this yields the well-known results that closed cages contain exactly twelve pentagons, a variable number of $F_6 = N/2 - 10$ hexagons and an even number of particles [33]. The snapshots suggest that a larger pucker angle χ yields smaller cages, and hence that F_6 and N are functions of χ, as will be discussed below. Henceforth two particles are regarded as 'bound' if their

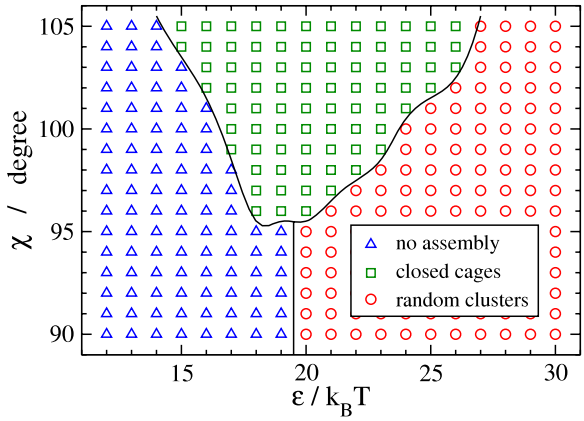

Figure 6. Phase diagram of the self-assembly behaviour against the leg–leg binding energy ϵ and the pucker angle χ of the legs. On the left-hand side the interactions are too weak to achieve clustering within $\sim 10^{10}$ MC steps, while the very 'sticky' particles on the right-hand side result in the rapid formation of many incomplete structures. Closed cages, i.e. clusters with a particle at every vertex and two anti-parallel legs along every edge, are produced at a high success rate in an approximately triangular intermediate region of parameter space, while they are rare in the other two regions.

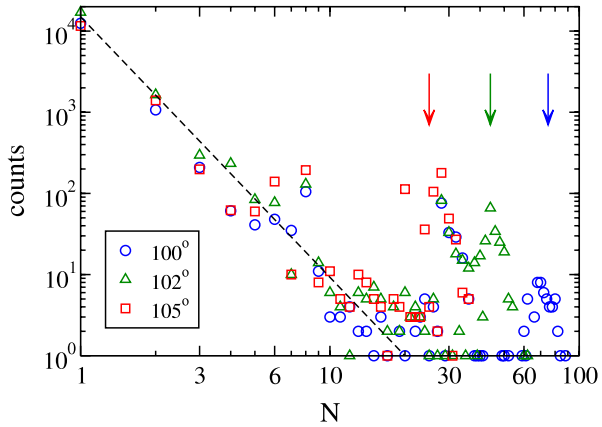

Figure 7. Distributions of cluster sizes after $\sim 10^{10}$ MC steps with $\sim$30 000 particles, for three different pucker angles and $\epsilon = 20 k_B T$. The number of small multimers on the left decays quickly with the cluster size, with the dashed line serving as a guide to the eye. The three peaks on the right, marked with arrows, clearly indicate that the average size of the self-assembled cages decreases with increasing pucker angle.

interaction energy is more negative than $-2k_B T$ and a cage is defined as 'closed' if every particle in the cluster is bound to three neighbours.

The phase diagram in figure 6 shows the self-assembly behaviour of the particles as a function of the two most important model parameters, namely the binding energy ϵ and the pucker angle χ. For the low concentrations studied here, a binding energy of at least $\sim 15 k_B T$ is required for the formation of multiple-particle aggregates. A low binding energy reduces the probability of several particles to unite into a critical nucleus, as well as promotes the probability of aggregates (including pre-formed closed cages) to fall apart into monomers. A high binding energy enhances the aggregation of particles, but too strong interactions make the particles cluster so rapidly that numerous cage fragments are being formed until the free monomers are depleted; the subsequent pairing of the diffusing clusters proceeds very slowly, with the lack of internal rearrangements making the combination of several erratic fragments into a closed cage a highly improbable event. Closed cages are only formed with a high success rate at intermediate binding energies, $\epsilon \sim 20 k_B T$, and for pucker angles exceeding $\sim 95°$. The latter limit arises because closed cages rapidly increase in size with decreasing χ, and thus exceed the available number of particles for χ approaching 90°.

The growth of closed cages from initially randomly distributed particles, under conducive conditions, proceeds by a nucleation-and-growth mechanism. This is clear from watching movies of the simulations, where small aggregates of a few particles are continuously being assembled and disassembled. Occasionally one of these aggregates crosses the transition state to start a steady growth to a stable cage structure. At the current density and binding energies, a newly created box of 1000 particles typically develops several simultaneously growing clusters. Later on in the simulations, when a considerable fraction of particles has been bound in clusters, the chances of forming a next stable cluster are considerably reduced. This drastic slowing down of the nucleation rate makes it difficult to assess whether an equilibrium state has been attained by the time the simulations are terminated.

Typical distributions of cluster sizes at the end of the simulations are shown in figure 7 for three puckers at $\epsilon = 20 k_B T$. For every χ, just over half of the $\sim$30 000 particles are still unbound after $(1 - 10) \times 10^{10}$ MC steps. There are numerous small aggregates, dimers, trimers etcetera, but their counts decline rapidly with increasing multimer size and drop to one or zero beyond $N \sim 10$. The plot contains several pronounced peaks at larger cluster sizes, corresponding to near-spherical cages, with the centre of the peak shifting to larger cages with decreasing pucker angle. Closer inspection of these clusters, both visually and by analysis software, reveals that many of them are completed polyhedrons, with a particle centred at every vertex and two anti-parallel legs along every vertex; these closed clusters are singled out in figure 8. The largest non-closed clusters are typically one or two particles short of a closed polyhedral structure. Occasionally we observe partly intertwined cages, with the legs of one cage sticking through the open faces of the other cage and vice versa, and in one simulation a cage was found to completely envelope another cluster; these few exceptional structures have for clarity been excluded from the distributions in figure 7.

The size distributions of closed cages at five different puckers are presented in figure 8. With decreasing pucker angle, the distribution shifts to higher N and becomes broader, indicating that there is more variation in the assembled structures. At the highest pucker angle, $\chi = 110°$, only icosahedral clusters are produced (not shown in figure 8). For $\chi = 105°$ eight different cluster sizes are observed; the gap at $N = 22$ arises because it is impossible to create a closed cage with twelve pentagons and one hexagon. A visual inspection

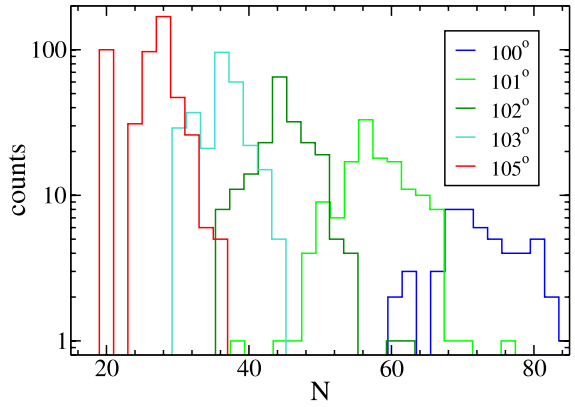

Figure 8. Distributions of cluster sizes for closed cages that self-assembled from ∼30 000 particles in $\sim 10^{10}$ MC steps with $\epsilon = 20k_BT$ at five different pucker angles.

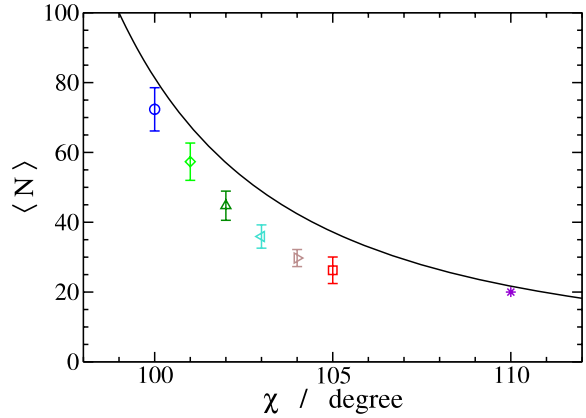

Figure 9. Average size $\langle N \rangle$ of self-assembled closed cages, as a function of the pucker angle χ, with the error bars denoting standard deviations. The solid line, representing equation (8), follows from a purely geometrical consideration.

of a dozen $N = 28$ cages indicates that most of these cages have identical structures, i.e. configurations of hexagonal and pentagonal faces. The questions why a small subset of cages is much more abundant than the numerous other permitted cage structures, and what property sets these few cages apart from the rest, are currently under investigation. One possibility is that these few structures are thermodynamically more stable, but it is equally well conceivable that the kinetics of the growth process directs the assembly towards certain preferred structures. Independent of cage size, the average potential energy of the closed cages typically lies between 80% and 90% of the maximum attainable energy of $3\epsilon N/2$. Figure 8 furthermore shows that the total number of assembled cages decreases with increasing pucker angle. This decline may have many causes, including variations in cage stability and chemical potential with χ, a similar reduction in free monomer concentration is achieved by fewer large cages, different assembly rates, and an increased probability of defects in larger cages.

The average size of the closed cages is plotted in figure 9 against the pucker angle. The graph shows that the average cluster size steadily increases with decreasing pucker angle, starting with $N = 20$ for dodecahedra grown at a pucker angle of 110° and ending at $\langle N \rangle \sim 70$ for $\chi = 100°$. Even larger aggregates were formed in simulations with lower pucker angles, like conical and flat fragments of 'chicken-wire' at 95° and 90° respectively, see figure 5, but these did not form closed cages and are therefore excluded from the plot. Clearly, $\langle N \rangle$ will diverge for χ approaching 90°. An approximate theoretical description of the data can be derived using Descartes' formula, stating that the total angle deficit of a polyhedron is 4π [33]. The angle deficit of a vertex may be idealized as 2π minus three times the leg–leg angle, which is readily calculated from the pucker angle. Multiplication by the number of vertices then yields

$$N\left[2\pi - 3\text{acos}\left(\frac{3\cos^2\chi - 1}{2}\right)\right] = 4\pi. \qquad (8)$$

This approximation for the average number of particles per closed cage lies systematically higher than the simulation results, see figure 9. In this simple derivation we have ignored that the two legs along an edge are often not perfectly aligned and, consequently, that the faces of the cage need not be flat, while thermodynamic and kinetic properties of the assembly process may also contribute to the observed difference.

5. Discussion and conclusions

The simulations have clearly highlighted the two key ingredients for the self-assembly of three-legged puckered particles into closed cages, namely anti-parallel binding of the legs and a dihedral preference around the leg–leg bond. The condition of anti-parallel binding of legs can be met in an experimental particle, and probably is met in the clathrin triskelion, by employing a series of weak interaction sites along the legs. One readily envisages how a non-equidistant distribution of sites, possibly in combination with discriminating interaction sites, makes anti-parallel binding highly favourable over parallel binding. The desired polarity of the legs may be created by concentrating the interaction sites on one face of the legs, as this will force the legs to face each other with their 'patchy' sides and thereby induces the dihedral preference that proved so crucial in the formation of cages. The concentration of binding sites on one side of the leg simultaneously creates a situation in which two bound legs are likely to prevent, by shielding the patchy sides with their excluded volumes, a third leg from attaching to the pair. We thus see how a collaborative effort by many short-range forces like Van der Waals and hydrophobic interactions, possibly assisted by hydrogen bonding and salt bridges, suffices to promote the self-assembly of large clathrin cages. In fact, long-range interactions such as magnetic or electric dipoles (with their dipolar moments arranged to create the desired torsional potential) are probably disadvantageous for the assembly process, since they will attract too many legs to the vertices and thereby form erratic aggregates, as observed in the early simulation model based on attractive interactions only.

Whether the polarity of the legs distinguishes between front and back of the legs, as in the current simulations, or between top and bottom, is of little consequence. Simulations employing the latter choice (and a potential g that promotes parallel polarities) indicate that a different definition of the polarity vector $\hat{\mathbf{m}}_{ib}$ will slightly alter the distribution of angles between neighbouring legs, but does not to affect the self-assembly characteristics as long as a sufficiently strong dihedral interaction orients the legs (and thereby the normals) of neighbouring particles in a lattice. The assembly process appears not to be sensitive to the exact form of the dihedral potential either, as cages are still being formed after replacing the smooth dihedral definition of $g(\mu)$ in equation (7) by the step function $g(\mu) = (\mu - |\mu|)/(2|\mu|)$.

A subtle difference between the simulated particles and real particles results from the absence of excluded volume interactions in the former: in the simulations the legs merely have to be oriented properly to attain the strongest interaction, while in the above scenario of legs with localized interaction sites the legs have to be oriented appropriately as well as being positioned properly relatively to each other. A related problem arises from the requested ability of the particles to form closed loops with both even (six) and odd (five) particles, with the latter obviously disallowing alternating sequences. We note that these conditions pose design restrictions on the locations of the interaction sites along the legs, and possibly also necessitate slightly curved legs to avoid packing problems due to overlapping legs.

The formation of closed or nearly closed cages proceeds remarkably efficiently for conducive particle parameters from a triangular area of parameter space. It is intuitively clear that growing a cage by randomly combining twelve pentamers and an even number of hexamers is very unlikely to result in a closed cage. The simulations indicate, however, that the random addition of triskelia to the unpaired legs of a growing cluster—within the restrictions posed by the potential—is remarkably likely to result in a closed cage. It appears as if the potential provides an innate guiding mechanism that directs the growing aggregates through a 'funnel' to the (nearly) closed cages, with the pucker determining the size of the cages. Both thermodynamic (energetic or entropic preference) and kinetic properties (availability of binding sites, ease of closing a five ring) may contribute to 'steering' the assembly process. A detailed discussion of which cluster structures are permitted, and what feature(s) makes them stand out from the long list of permitted polyhedra, lies beyond the scope of the current study. To this effect, we are currently working on the automated recognition of the over a thousand self-assembled cage structures; their analysis and a comparison with theoretical predictions will be presented in a forthcoming paper.

A peculiar property of Monte Carlo simulations that surfaced in our simulations, as well as in other MC patchy particle simulations [34], is that clusters appear to move much slower than monomers. Some slowing down is expected, since the diffusion coefficient of an N-mer scales inversely proportional with N, but the slowing down in the MC simulations is much stronger. We believe that this is not a problem in the current study, because the rate of aggregation at a slowly moving cluster is largely determined by the highly mobile loose monomers. Slowing down might, however, introduce artefacts under some conditions, and more advanced cluster moves are under development to compensate these effects [35, 36].

We end by noting that the optimum binding energies in this study, $\epsilon \sim 20k_{\mathrm{B}}T$, are somewhat larger than the values of $\sim 7k_{\mathrm{B}}T$ reported by Wilber *et al* [25] and $(12 - 18)k_{\mathrm{B}}T$ by Hagan and Chandler [24] in their simulations of nearly-spherical patchy particles forming icosahedral cages with $N = 12$ and $N = 60$, respectively. This difference probably reflects the lower number density in our simulations (in this study $\rho = 10^{-3}\sigma^{-3}$, versus $0.15\sigma^{-3}$ and $(0.04 - 0.5)\sigma^{-3}$ in the two preceding studies, though this comparison is obviously hampered by the distinct particles used in the three simulations), which necessitates a correspondingly higher binding energy to combat the larger loss of translational entropy by a particle upon binding to a cluster. It would be advantageous for comparison purposes, across simulations and with experiments, to eliminate the concentration dependency by determining the equilibrium constant of the $NX \rightleftharpoons X_N$ assembly process, which is closely related to the free energy difference between free monomers and assembled clusters. A direct calculation of the free energy difference by dedicated methods [24, 25] would also provide an independent estimate of a patchy particle analogue to the familiar critical micelle concentration [7], and a measure for the degree of equilibration that has been achieved in the simulations—since the dynamics of cluster assembly and disassembly are fairly slow, it is currently difficult to judge from the simulations whether equilibrium has actually be attained. Establishing these free energies for our three-legged particles is inhibitively computer time consuming due to the time needed to establish the free energy of a single cage structure and the large number of self-assembled structures observed in our simulations.

References

[1] Glotzer S C 2007 Anisotropy of building blocks and their assembly into complex structures *Nat. Mater.* **6** 557–62

[2] Zhang Z, Keys A S, Chen T and Glotzer S C 2005 Self-assembly of patchy particles into diamond structure through molecular mimicry *Langmuir* **21** 11547–51

[3] Rothemund P W K 2006 Folding DNA to create nanoscale shapes and patterns *Nature* **440** 297–302

[4] Douglas S M, Dietz H, Liedl T, Högberg B, Graf F and Shih W M 2009 Self-assembly of DNA into nanoscale three-dimensional shapes *Nature* **459** 414–8

[5] Zandi R, Reguera D, Bruinsma R F, Gelbart W M and Rudnick J 2004 Origin of icosahedral symmetry in viruses *Proc. Natl Acad. Sci. USA* **101** 15556–60

[6] Zandi R, van der Schoot P, Reguera D, Kegel W and Reiss H 2006 Classical nucleation theory of virus capsids *Biophys. J.* **90** 1939–48

[7] Gelbart W M, Ben-Shaul A and Roux D (ed) 1994 *Micelles, Membranes, Microemulsions and Monolayers* (Berlin: Springer)

[8] Boal D 2002 *Mechanics of the Cell* (Cambridge: Cambridge University Press)

[9] Brodsky F M, Chen C-Y, Knuehl C, Towler M C and Wakeham D E 2001 Biological basket weaving: formation and function of clathrin-coated vesicles *Ann. Rev. Cell Dev. Biol.* **17** 517–68

[10] Young A 2007 Structural insights into the clathrin coat *Sem. Cell Dev. Biol.* **18** 448–58
[11] Schmid E M and McMahon H T 2007 Integrating molecular and network biology to decode endocytosis *Nature* **448** 883–8
[12] Rappoport J Z 2008 Focusing on clathrin-mediated endocytosis *Biochem. J.* **412** 415–23
[13] Wakeham D E, Chen C-Y, Greene B, Hwang P K and Brodsky F M 2003 Clathrin self-assembly involves coordinated weak interactions favorable for cellular regulation *EMBO J.* **22** 4980–90
[14] Fotin A, Cheng Y, Sliz P, Grigorieff N, Harrison S C and Kirchhausen T 2004 Molecular model for a complete clathrin lattice from electron cryomicroscopy *Nature* **432** 573–9
[15] Kroto H W, Heath J R, O'Brien S C, Curl R F and Smalley R E 1985 C_{60}: Buckminsterfullerene *Nature* **318** 162–3
[16] Rapaport D C, Johnson J E and Skolnick J 1999 Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation *Comput. Phys. Commun.* **121** 231–5
[17] Ford M H, Auerbach S M and Monson P A 2004 On the mechanical properties and phase behavior of silica: a simple model based on low coordination and strong association *J. Chem. Phys.* **121** 8415–22
[18] Rapaport D C 2004 Self-asssembly of polyhedral shells: a molecular dynamics study *Phys. Rev.* E **70** 051905
[19] Zhang Z and Glotzer S C 2004 Self-assembly of patchy particles *Nano Lett.* **4** 1407–13
[20] Van Workum K and Douglas J F 2006 Symmetry, equivalence, and molecular self-assembly *Phys. Rev.* E **73** 031502
[21] Nguyen H D, Reddy V S and Brooks C L III 2007 Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids *Nano Lett.* **7** 338–44
[22] Schwartz R, Shor P W, Prevelige P E Jr and Berger B 1998 Local rules simulations of the kinetics of virus capsid self-assembly *Biophys. J.* **75** 2626–36
[23] Kern N and Frenkel D 2003 Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction *J. Chem. Phys.* **118** 9882–9
[24] Hagan M F and Chandler D 2006 Dynamic pathways for viral capsid assembly *Biophys. J.* **91** 42–54
[25] Wilber A W, Doye J P K, Louis A A, Noya E G, Miller M A and Wong P 2007 Reversible self-assembly of patchy particles into monodisperse icosahedral clusters *J. Chem. Phys.* **127** 085106
[26] Noya E G, Vega C, Doye J P K and Louis A A 2007 Phase diagram of model anisotropic particles with octahedral symmetry *J. Chem. Phys.* **127** 053501
[27] Fantoni R, Gazzillo D, Giacometti A, Miller M A and Pastore G 2007 Patchy sticky hard spheres: analytical study and Monte Carlo simulations *J. Chem. Phys.* **127** 234507
[28] Huisman B A H, Bolhuis P G and Fasolino A 2008 Phase transition to bundles of flexible supramolecular polymers *Phys. Rev. Lett.* **100** 188301
[29] Tao Y-G, den Otter W K, Dhont J K G and Briels W J 2006 Isotropic–nematic spinodals of rigid long thin rodlike colloids by event-driven Brownian dynamics simulations *J. Chem. Phys.* **124** 134906
[30] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (Oxford: Oxford University Press)
[31] Frenkel D and Smit B 1996 *Understanding Molecular Simulation* (San Diego, CA: Academic)
[32] Humphrey W, Dalke A and Schulten K 1996 VMD: visual molecular dynamics *J. Mol. Graph.* **14** 33–8
[33] Richeson D S 2008 *Euler's Gem: The Polyhedron Formula and The Birth of Topology* (Princeton, NJ: Princeton University Press)
[34] Louis A A, private communication
[35] Whitelam S and Geissler P L 2007 Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles *J. Chem. Phys.* **127** 154101
[36] Bhattacharyay A and Troisi A 2008 Self-assembly of sparsely distributed molecules: an efficient cluster algorithm *Chem. Phys. Lett.* **458** 210–13