Modelling the self-assembly of virus capsids

Modelling the self-assembly of virus capsids

Iain G Johnston[1], Ard A Louis[1] and Jonathan P K Doye[2]

[1] Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, UK
[2] Physical and Theoretical Chemistry Laboratory, Department of Chemistry,
University of Oxford, South Parks Road, Oxford OX1 3QZ, UK

Abstract
We use computer simulations to study a model, first proposed by Wales (2005 *Phil. Trans. R. Soc.* A **363** 357), for the reversible and monodisperse self-assembly of simple icosahedral virus capsid structures. The success and efficiency of assembly as a function of thermodynamic and geometric factors can be qualitatively related to the potential energy landscape structure of the assembling system. Even though the model is strongly coarse-grained, it exhibits a number of features also observed in experiments, such as sigmoidal assembly dynamics, hysteresis in capsid formation and numerous kinetic traps. We also investigate the effect of macromolecular crowding on the assembly dynamics. Crowding agents generally reduce capsid yields at optimal conditions for non-crowded assembly, but may increase yields for parameter regimes away from the optimum. Finally, we generalize the model to a larger triangulation number $T = 3$, and observe assembly dynamics more complex than that seen for the original $T = 1$ model.

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

1. Introduction

One of the simplest examples of self-assembly in biology is that of the virus capsid. The protective protein coat surrounding the viral genetic material is assembled into its monodisperse form reversibly from a large number of quasi-identical subunits, or capsomers. In some cases, this can even occur *in vitro*, as famously first demonstrated for the tobacco mosaic virus where the capsids dissociated upon raising the pH of a solution, but reversibly reassembled into complete capsid structures upon the subsequent lowering the pH back to the initial conditions [2]. Although in nature the assembly of virus capsids can be more complex, with nucleic acids, scaffolding proteins, and other constituents playing a role, there are a good number of viruses where successful reversible *in vitro* assembly can occur from just the purified proteins. Well studied examples include icosahedral viruses such as the cowpea chlorotic mottle virus (CCMV) [3], the hepatitis B virus (HBV) [4] and the human papillomavirus (HPV) [5]. These experiments suggest that it is possible to encode all the necessary assembly information into the individual capsomers themselves. As such, virus self-assembly *in vitro* is a paradigmatic example of monodisperse self-assembly. A better understanding of the underlying physics

involved will not only lead to new biological insights, but may also stimulate novel applications in nanotechnology.

Viruses typically vary in size from about 2 to 200 nm in diameter and are extremely successful organisms. They can be found in a wide variety of environments and it has recently been estimated that they constitute a larger fraction of the total biomass on earth than eukaryotes (the family that includes animals and plants) [6]. About half of all virus families have icosahedral symmetry. As first pointed out by Caspar and Klug [7], the structure of these icosahedral virus capsids can be understood using the concept of quasi-equivalence, whereby identical proteins can occupy different location types within the capsid structure. Some capsomers group around axes of five-fold symmetry, and others around axes of six-fold symmetry, occupying geometrically distinct locations despite their structural similarity [8–10]. A simple geometric argument suggests that capsids can be made up of 12 pentagonal and $10(T - 1)$ hexagonal structural units, where the triangulation number T is restricted to numbers that can be expressed by $T = a^2 + ab + b^2$, with a and b non-negative integers (e.g. $T = 2$ is not possible, but $T = 1$ and 3 are). The smallest $T = 1$ viruses are made up of just 12 pentameric units. Many of these are satellite viruses that are parasitical to larger viruses, but others such as the alfalfa mosaic virus [11] are viable on

their own. More complex viruses have higher triangulation numbers, including CCMV ($T = 3$), HBV ($T = 4$) and HPV ($T = 7$) [11], and the largest to date is estimated to be an enormous $T \simeq 1000$ for the giant Mimivirus [12].

Experimental studies have observed the *in vitro* assembly of virus capsids using light scattering [13, 14] and electron microscopy [13, 15]. Several key features are commonly observed. Firstly, the yield of complete capsids is found to be a sigmoidal function of time, with an initial lag period during which no capsids are produced, followed by a period of rapid completion before the yield reaches a plateau. Secondly, under fast growth conditions such as high subunit concentrations or strong attractions, assembly is seen to be vulnerable to kinetic traps in which an excessive number of nuclei is formed. Each of these nuclei grow, consuming subunit monomers in the process and leading to a rapid depletion of the monomer population ('monomer starvation') and the eventual production of a large number of part-formed capsids. Also observed is the formation of 'monster particles', consisting of malformed clusters of part-formed capsids. Thirdly, hysteresis is also commonly found, where fully formed capsids remain stable in conditions under which they would not form from individual capsomers.

Modelling virus assembly at the atomistic level is prohibitively expensive. In a recent study, a fully atomistic simulation of a complete $T = 1$ icosahedral satellite tobacco mosaic was performed [16]. This simulation involved over a million individual atoms (mostly solvent), and so could only sample about 50 ns of time, whereas dynamic assembly occurs on much longer timescales (up to seconds).

Many studies of virus assembly have thus, by necessity, treated the process by using strongly coarse-grained models. Important early work was done by Zlotnick and *et al* [5, 17], who employed kinetic equations that measure the flux between populations of different sized assemblies. The spatial location of the different species is averaged over, and in most of the work, the sets of possible reactions and intermediates was limited to the addition of single capsid proteins to their equilibrium position in capsomers, thus ignoring many potential intermediate states. Nevertheless, these kinetic studies reproduced important experimentally observed features such as monomer starvation and hysteresis, and showed that relatively weak association energies are sufficient for capsid assembly. The question of how important the approximation of neglecting all but the most stable intermediates in the assembly pathway is still under active debate [18–20].

Direct simulations of model coarse-grained protein assemblies have the advantage that spatial fluctuations and a much wider range of intermediate states are naturally included. Some of the first were performed by Rapaport *et al* [21, 22] who used molecular dynamics (MD) simulations of triangular and trapezoidal units that assemble into capsids. However, these simulations suffered from drawbacks, such as the use of unrealistic ballistic dynamics, and not always satisfying detailed balance. In a recent paper, Rapaport performed fully reversible simulations where complete icosahedra assembled from 20 triangular particles [23]. A background fluid was used in order to approximate the diffusive dynamics of proteins in

solution, and detailed balance was also satisfied. There is now a considerable body of work using MD, Brownian dynamics or Monte Carlo (MC) simulations of coarse-grained particles to model the self-assembly of viruses [19, 24–28] and other objects that self-assemble into monodisperse clusters [29–33]. These studies exhibit a number of similar trends. For example, to achieve self-assembly, the temperature must be low enough that the target structure is thermodynamically stable, but not so low that incorrectly bonded particles cannot separate. Similarly, the design parameters must be specific enough to favour the target structure over alternative structures, but not so constrained as to hinder kinetic accessibility of the desired structure [29]. Given that these basic trends are seen by such a wide variety of models, it suggests that they are robust features of self-assembly that will be relevant to the physical case of virus self-assembly *in vitro*.

A different way of approaching the design problem for virus capsids, and other self-assembling systems, is to study the potential energy surface (PES), a high dimensional function that describes how the potential energy depends on the coordinates of the N particles in a system. It is closely related to the concept of an energy landscape or a free-energy landscape if temperature is included [34, 35]. This point of view has been particularly explored in the context of protein folding, which can also be viewed as a self-assembly phenomenon, where the particles (the amino acids) are connected together, rather than being free as they are in virus self-assembly. An important concept for protein folding is the idea of a 'folding funnel' [34, 36–38], that helps explain how a protein can overcome the Levinthal paradox [39], which states that it is impossible for a protein to find its folded state on a physical timescale by a completely blind search because the number of states accessible to a typical protein is astronomically large. Instead, the 'funnel' topology of the PES helps guide the system through a directed search towards the free-energy minimum. By analogy, one might expect that the energy landscape of a self-assembling system must also show a funnel-like topology if the system is to assemble. This funnel topology, also referred to as a 'palm-tree' PES, is predicted to be a ubiquitous feature of good 'structure-seeking' systems [35]. Indeed, we have recently calculated free-energy landscapes of model patchy particles, and shown that this feature can help rationalize the dependence of self-assembly yields on design parameters [27, 33].

In a recent paper, Wales [1] proposed a model for the assembly of $T = 1$ viruses based on a set of 12 pentagonal units, which has since been recently extended [40] to model $T = 3$ capsids. In [1], the PES for a single connected 12-mer was characterized, and stationary states were used to analyse the topology of the landscape for a number of different design parameters. Rugged, glassy landscapes and landscapes with insufficient funnelling were proposed to hinder assembly. Those landscapes that exhibit a more funnel-like topology were predicted to promote assembly. However, there are a few caveats to this picture. Firstly, the PES was calculated for a connected 12-mer, which on its own does not explain why 12-mers form rather than clusters of other sizes. However, in reference [40], an extended model was used to calculate the

PES for 24 particles, and shown to have separate funnels, one corresponding to two distinct icosahedral capsids and the other to a single 24-particle cluster, the former being more stable. More generally, the PES is a very high dimensional function, but the analysis of its topology is usually done by projecting it down to a much lower dimensional representation [35]. The robustness of these PES-based predictions is therefore an interesting subject of investigation.

To investigate the PES predictions, we perform MC computer simulations on the same model and study its assembly dynamics. We find good agreement with Wales' predictions, giving further evidence that the energy landscape picture can help rationalize how nature achieves the design of individual particles that can self-assemble into well defined monodisperse shapes [35]. Even though this model is strongly coarse-grained, we observe behaviour that is also seen in experiments, including sigmoidal assembly dynamics, hysteresis in capsid formation, and a variety of different kinetic traps.

We also extend the model in two ways. We first add crowding agents to model the fact that when viruses assemble *in vivo*, they do so in the densely packed environment of the cell. The crowding lowers the assembly efficiency compared to the uncrowded case for parameters near the optimal assembly, but can increase efficiency in regions where the uncrowded virus assembly proceeds less well.

We next extend the model to $T = 3$ viruses by including 20 hexagonal units per capsid, and find a region of parameter space where assembly is successful. We also investigate a number of different kinetic traps and pathways to assembly not observed for the $T = 1$ model. In contrast to the simpler $T = 1$ model, the assembly efficiency drops considerably when more than one capsid is simulated, in part because the pentagons and the hexagons must come together in the right numbers per capsid.

We proceed as follows: in section 2 we describe the model and the MC simulation method we use, and in section 3 we discuss simulation results for the $T = 1$ assembly dynamics. Section 4 includes the two extensions of our model, namely crowded assembly and the $T = 3$ virus model.

2. Methods

2.1. Model

We use the same model as Wales [1]. The capsomers are represented as rigid pentagonal pyramids with a distance r_b from the centre of the capsomer base to the basal vertices, and with an apex site at a height h above the centre of the base, as illustrated in figure 1. Representing a capsomer c_i by vectors corresponding to its base points, $\{\mathbf{p_i^1}, \ldots, \mathbf{p_i^5}\}$, and its apex $\mathbf{a_i}$, the potential between two capsomers c_i and c_j is:

$$V(c_i, c_j) = V_{\text{apex}}(|\mathbf{a_i} - \mathbf{a_j}|) + \sum_{u=1}^{5} \sum_{v=1}^{5} V_{\text{M}}(|\mathbf{p_i^u} - \mathbf{p_j^v}|) \quad (1)$$

where $V_{\text{M}}(r)$ is a Morse potential of the form:

$$V_{\text{M}}(r_{ij}) = \epsilon \left(e^{\rho(1 - \frac{r_{ij}}{r_e})} - 2 \right) e^{\rho(1 - \frac{r_{ij}}{r_e})} \quad (2)$$
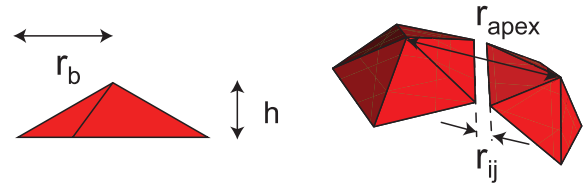
Figure 1. The rigid pentagonal pyramid capsomer units have a height of h and a distance between the base centre and the vertices of r_b. Also shown here is the distance r_{apex} between two apices and the distance r_{ij} between two basal vertices on different capsomers.

and $V_{\text{apex}}(r)$ is a purely repulsive interaction of the form:

$$V_{\text{apex}}(r) = \epsilon_{\text{R}} \left(\frac{\sigma}{r_{\text{apex}}} \right)^{12}. \quad (3)$$

Here ϵ defines the unit energy (from which reduced temperature T is derived), and r_b, the distance between the centre of a capsomer's base and one of its vertices, defines unit distance. r_e, the length scale of the Morse potential, is set to $0.2r_b$. The range of the interaction is set to $\rho = 0.65r_b$, and the strength of the repulsive term is set to $\epsilon_{\text{R}} = \frac{\epsilon}{2}$. Capsomers can have variable height h, and unless otherwise noted, h is set to $r_b/2$. The apex–apex repulsion length scale is set to $\sigma = 2.1r_b$, the distance between two apices in a complete capsid when $h = 0.5r_b$ [1], so that $V_{\text{apex}}(r) = \epsilon_{\text{R}}$ when the capsid is fully formed. The attractions between the vertices allow the capsomers to bond, and the repulsive interaction between the apex sites sets the curvature of the capsid, see figure 2 for an illustration of these potentials. We note that the torsional flexibility allowed by the repulsive interaction, illustrated in figure 2(b), is greater than that expected for the protein–protein bonding found in nature, which typically involves several interacting sites and hence more angular specificity. Generic self-assembly properties, and the role of the PES, can be freely investigated with this system, but this increased flexibility should be kept in mind when considering biological virus assembly.

2.2. Simulations

To perform the simulations we employed a standard metropolis Monte Carlo scheme with only local translation and rotation moves of the individual capsomers. The small scale and random nature of the steps applied generate diffusive motion similar to that expected for proteins in solution [41, 42]. The real dynamics of aggregating particles in solution are more complicated [43], and hydrodynamic effects, which are neglected in Brownian dynamics, may also play a role [44]. However, local move MC should be adequate to capture overall trends in a diffusion limited system. Time in the MC simulations employed is discrete, with each MC step involving a single attempted move, and an MC cycle involving N steps. Translation and rotation moves are chosen stochastically, each with a 50% probability. Translational moves involve an entire pentagonal capsomer being translated by a randomly chosen vector with magnitude less than a chosen cutoff magnitude. Rotational moves involve a capsomer being rotated by an angle
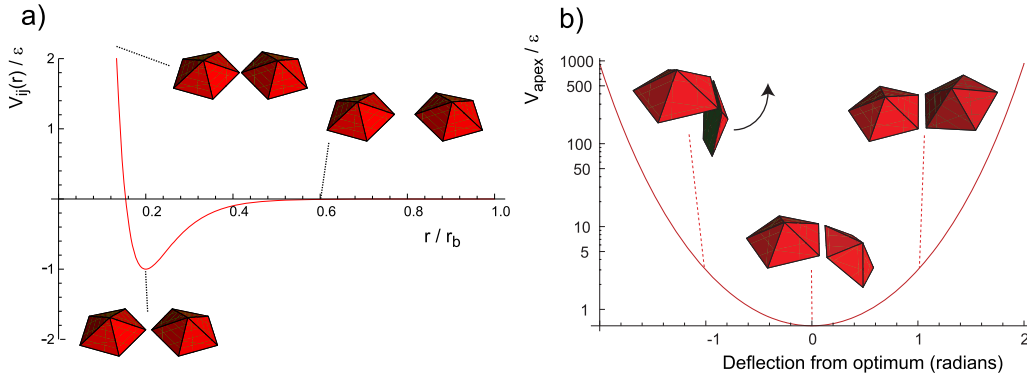
Figure 2. (left) Morse interaction $V_{ij}(r)$ between base points on model capsomers. (right) The repulsive potential $V_{apex}(r)$ between capsomer apices helps set the curvature of the fully formed capsids. Note that capsomer illustrations are not to same scale as graphs.

α, less than or equal to a chosen cutoff angle, through the centre of the capsomer's base, about a randomly chosen axis. The cutoffs in step magnitude were chosen to give a move acceptance ratio of approximately 50%. Due to the coarse-grained and conceptual nature of this model, we did not attempt to adjust move size and frequency to reflect quantitatively accurate ratios of rotational to translational diffusion.

The simulations were performed in a cubic simulation box in continuous space with periodic boundary conditions. The number N_5 of pentagonal capsomers was varied between 12 and 120. To initialize the system, 10^4 MC cycles (where a cycle is N_5 steps) were performed at $T^* = 10\epsilon k_B^{-1}$. Assembly was then monitored for runs up to 5×10^5 MC cycles unless otherwise stated.

To analyse the clusters of bonded capsomers we employed the following protocol. Two capsomers c_i and c_j are considered to be bonded if $V(c_i, c_j) < -\epsilon$. Since the potential energy of two capsomers in a perfectly bonded configuration is $-2\epsilon + \epsilon_R$ (terms respectively from the base-point bonds and apex repulsion), equal to $\frac{3}{2}\epsilon$, our bonding definition allows a degree of thermal fluctuation while ensuring that capsomers are still bound. A cluster is then a set of capsomers where each capsomer in the set is reachable from any other by following a series of bonds. We define a cluster size order parameter, C, as the size of the largest cluster present in a simulation as well as a geometric capsid order parameter, Q, defined as:

$$Q = \frac{1}{N_5}b_5, \quad (4)$$

where b_5 is the number of capsomers in the simulation bonded to exactly five other capsomers with bond angles consistent with that of a fully formed capsid structure. This geometric order parameter is useful to distinguish the correct structure from those with inverted capsomers, predicted by Wales [1] to represent significant kinetic traps in this model.

3. Results for $T = 1$ capsid assembly

3.1. Assembly yields with density and temperature

For a fixed height $h = 0.5r_b$, we simulated $N_5 = 120$ capsomers and studied how the assembly yield, defined as the
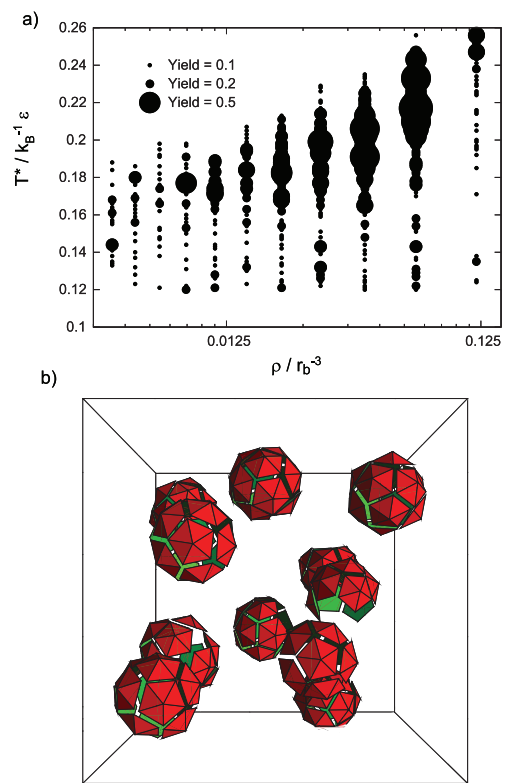
Figure 3. (a) Assembly yield with number density ρ and reduced temperature T. (b) Snapshot of capsid assembly at $T = 0.22\epsilon k_B^{-1}$, $\rho = 5.5 \times 10^{-2}r_b^{-3}$ (packing fraction $\phi \simeq 0.27$), showing a 60% yield of capsids. In this case there are 11 clusters, because a number of capsids are only partially formed and there are no more monomers left to complete the assembly (the monomer starvation trap).

fraction of fully formed capsids (maximum of 10), varies as a function of temperature and number density $\rho = N_5/L^3$, where L is the length of one side of the cubic simulation box (periodic boundary conditions are applied). The results are shown in figure 3. We observe firstly that for a fixed density, the region of optimal assembly is bounded from above and below in temperature. If the temperature drops too low, then the attractions are too strong and mis-bonded capsomers cannot
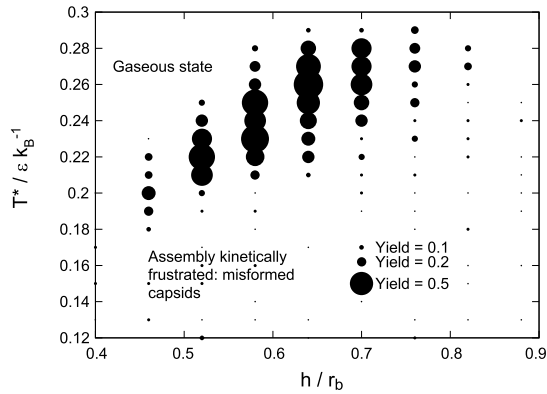
Figure 4. Assembly yield with capsomer height h. As predicted by Wales [1], the optimal assembly occurs at intermediate values of h where the PES is predicted to show 'funnel'-like features.

Figure 5. Low-temperature kinetic traps in the formation of a $T = 1$ capsid. (a) $h = 0.4r_b$: clustering into amorphous structure. (b) $h = 0.75r_b$: kinetic traps often involve one or more capsomers in the incorrect geometry, here inverted, as predicted by Wales [1].

dissociate and reassemble in correct configurations. If the temperature is too high, the attractions are not strong enough to ensure bonding, and the high entropy disordered state is favoured. These results mirror those found by a number of other investigators [24, 27, 29].

Similarly, for a fixed temperature, there is also a window of densities for which optimal yields are obtained. As discussed for example in the work of Hagan and Chandler [24], this trend is due to a tradeoff between having many subunit collisions with increasing density, and avoiding kinetic traps that create amorphous structures at higher densities. Indeed we find many partially formed shells at low densities, and amorphous bonded structures at higher densities. The optimal assembly region is around $\rho \simeq 5.5 \times 10^{-2}r_b^{-3}$, at $T \simeq 0.22\epsilon k_B^{-1}$. We define $\rho^* = 5.5 \times 10^{-2}r_b^{-3}$ for use in further simulations. To express this as a packing fraction, we approximate the volume occupied by each capsomer as a sphere of diameter σ, the length scale of the repulsive interaction. Using this measure, ρ^* corresponds to a capsomer packing fraction of $\phi = \pi\rho^*\sigma^3/6 \simeq 0.27$.

The yield of complete capsids is observed to be sigmoidal with time, with a brief period of zero yield during which nucleation occurs, followed by a rise in yield as capsids begin to form, plateauing as monomers are depleted and no further assembly can occur.

3.2. Assembly yield with capsomer height

Wales calculated the PES for different values of the capsomer height h [1]. For $h = 0.5r_b$, he predicted a funnel-like topology that would facilitate assembly. For a smaller capsomer height $h = 0.35r_b$, the fully formed icosahedral capsid is still the global minimum, but Wales predicts that the potential energy gradient and thus the driving force towards the formation of closed shells is diminished. Furthermore, he predicts a kinetic trap made up of loosely packed capsomers. Similarly, for larger capsomer height $h = 0.75r_b$, he predicts kinetic traps where capsomers join the capsid structure with their apices pointing inwards. Again, assembly is predicted to be hindered by these competing low-lying energy states. We
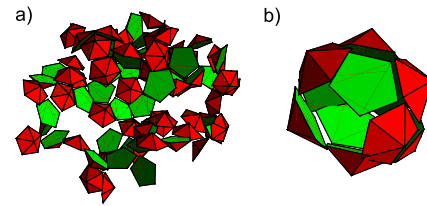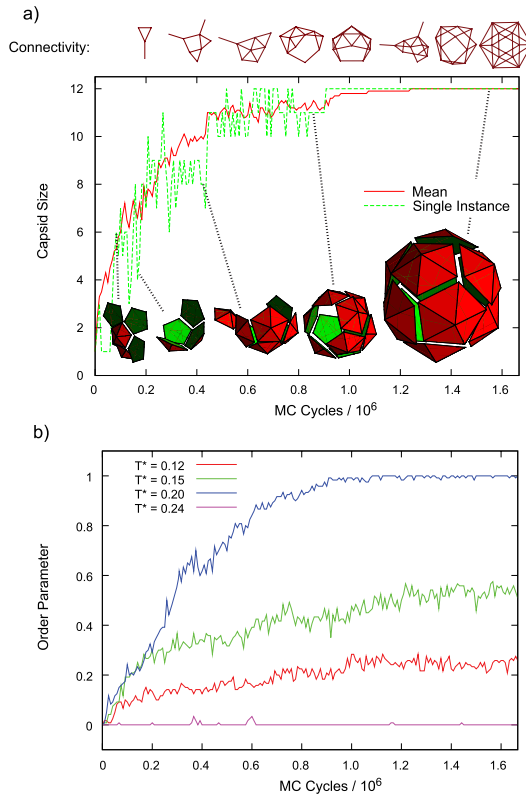
note that these trapping structures, due to the aforementioned differences between our model interaction potential and realistic protein–protein interactions, may not correspond exactly to the trapping structures found in biological virus assembly. However, the presence of trapping structures, and a region of parameter space in which they can be avoided, is of general interest in the study of self-assembling systems.

We simulated the system at a fixed density ρ^*, but for different capsomer heights and temperatures. The results are shown in figures 4 and 5. In agreement with Wales, we find that optimal assembly occurs at an intermediate h, with an optimum closer to $h = 0.6r_b$, and very low yields for $h < 0.35r_b$ or $h > 0.75r_b$. As illustrated in figure 5, for the larger h we observe kinetic traps characterized by shells with inverted capsomers and for low h we find kinetic traps where the system forms an extended amorphous structure.

3.3. Assembly mechanisms and thermodynamics for a single capsid

To study the assembly mechanism of a capsid in more detail, simulations were carried out for 12 capsomers at an effective of density ρ^* and for different temperatures T. The cluster size order parameter C and the geometric capsid order parameter Q were averaged over ten independent simulations.

We also monitored the geometry of intermediate structures on the route to full assembly. Many observed intermediates differ from the most stable structures utilized by rate equation approaches. Here figure 6 depicts the connectivity graphs of selected intermediate structures. Nodes represent capsomers in the largest cluster present and edges representing bonds between capsomers. The connectivity graphs of growing clusters often indicate structures with several 'leaves' (capsomers with only one bond to the rest of the cluster) whereas the most stable structure for a given size necessarily minimizes the number of leaves. In addition, intermediates exist where some capsomers are bonded to a cluster in geometrically incorrect positions, for example, inverted or differently angled. The very commonplace presence of these structures in assembly paths suggests that the subset used in rate equation studies may omit important detail about the assembly process [19, 20, 45]. On the other hand, the rather extreme nature of some structures, for example, those involving inverted capsomers, is a consequence of the aforementioned flexibility in the model interactions, and such extreme structures may not be biologically realistic.

Figure 7. Heat capacity from umbrella sampling, $N_5 = 120$, $\rho = \rho^*$, $h = 0.5r_b$. Upper axis shows temperature scaled such that the peak from simulation data matches the peak of an experimental heat capacity curve for a $T = 1$ parvovirus capsid [47]. Our model displays a full-width at half-maximum (FWHM) on this scale of approximately 7 °C, broader than the experimental FWHM of approximately 2.5 °C.

Figure 6. (a) Cluster size order parameter C as a function of simulation time during assembly at $T = 0.2\epsilon k_B^{-1}$. The cluster size averaged over ten simulations is given by the solid line and a typical single simulation is denoted by the dashed line. Also shown are snapshots and connectivity graphs of the largest cluster at intervals throughout assembly. (b) Geometric order parameter Q with time for different T.
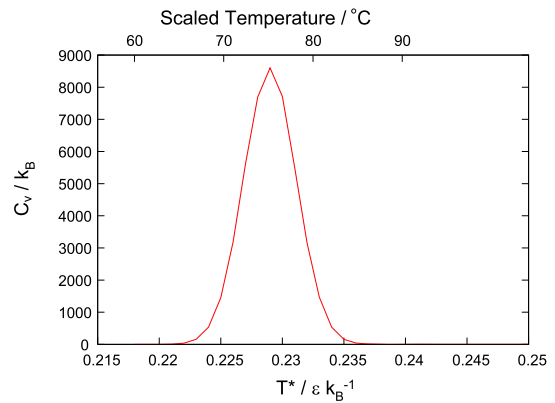
We also observed hysteresis: if a capsid was fully formed, and subsequently the temperature is raised, the temperature at which it breaks up is higher than the temperature at which it forms spontaneously from individual capsids. For simulations with the local MC moves, the difference between the superheated and undercooled temperatures were measured to be about $\Delta T = 0.17\epsilon k_B^{-1}$.

To fully sample the thermodynamics of the capsid formation, we also performed umbrella sampling MC simulations [46], with the total energy term adjusted by a term $V'(Q)$ dependent on the order parameter. This allows the system to sample extensively over the full range of Q. From this the melting point can be determined while the heat capacity C_v follows from fluctuations in energy. We observe a peak in C_v at the melting point of the capsid structure (see figure 7), and the area under the peak is approximately equal to the energy of the fully formed structure. However, the heat capacity peak is somewhat broader than that observed in biological studies [47, 48].

4. Crowding agents and models of $T = 3$ capsid assembly

4.1. Assembly yield with crowding agents

The focus of this paper has so far been on *in vitro* self-assembly of virus capsids. As a first step towards modelling

the biologically more relevant case of *in vivo* assembly, we introduced crowding agents into the simulation. In biology, the cytoplasm of a cell is typically filled with a significant volume fraction of proteins and other biomolecules [49, 50], and these are expected to affect the kinetics of assembly. Experimental studies have found the presence of crowding agents to facilitate *in vitro* assembly of HIV-1 capsids [51] and to stabilize the native states of folded proteins [52]. Two important competing effects of crowding on reaction rates have been noted in the literature [53, 54]. Firstly, crowding increases the effective local concentration of reactants, which generally increases reaction rates. Secondly, the crowding agents can also slow down the diffusion of the reactants, which can lower reaction rates, especially if they are diffusion limited. Thus the overall effect of crowding on reaction rates may arise from a complex interplay of different effects.

We model the crowding by introducing soft repulsive spheres interacting with the potential

$$V_{\text{crowd}}(r) = \epsilon \left(\frac{\sigma}{r} \right)^{12}, \qquad (5)$$

both between their centres and with the apices of the pentagonal pyramids. Simulations were run with $N_5 = 120$ capsomers at $\rho = \rho^*$. Assembly yields were measured for different crowding agent densities ρ_{crowd} and for different temperatures T. The simulation time was lengthened to 1.5×10^6 MC cycles, three times longer than the bare simulations, to take into account the fact that the crowding agents are expected to slow down the diffusion of the capsomers. Figure 8 shows the results from these experiments, in which increased crowding generally lowers the assembly yields. We attribute this change to an observed rapid decrease in the average diffusivity with the addition of crowding agents. Moreover, it should be kept in mind that these simulations were performed at $\rho = \rho^*$, the optimal density for assembly without crowders, so it is perhaps not surprising that the assembly yields generally decrease with the addition of crowding agents. However, we

a)

b)
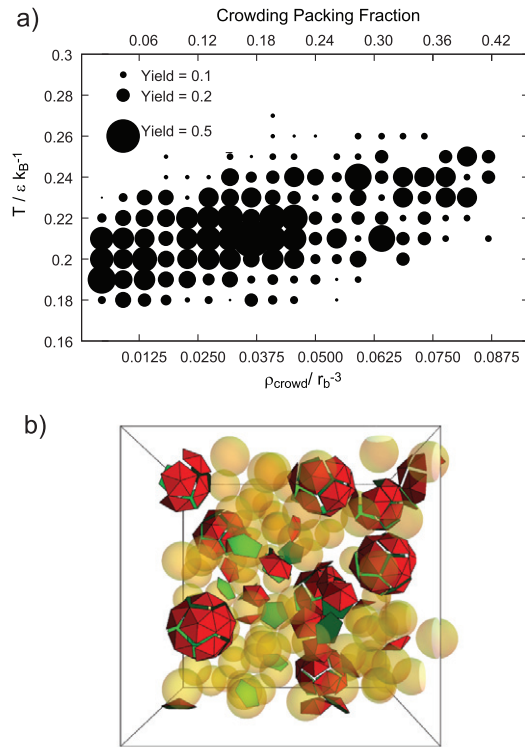
Figure 8. (a) Assembly yield with number density ρ_{crowd} of crowding agents and temperature T at a fixed capsomer density $\rho = \rho^*$. (b) Snapshot of assembly with $\rho_{crowd} = 4.1 \times 10^{-2} r_b^{-3}$ crowding agents (yellow) at $T = 0.22\epsilon k_B^{-1}$, showing several fully formed capsids.

also observe that crowding can increase yields at some higher temperatures. For example, we observe non-zero yields at $T = 0.27\epsilon k_B^{-1}$ at some of the higher crowding densities, while the yield is essentially zero there without crowding agents. More generally, the optimal assembly temperature shifts upward with increasing packing fraction of the crowding agents, mirroring the effect shown in figure 3 where the optimum temperature shifts upwards with increasing capsomer density. We thus attribute this effect to the decrease in free volume with increasing capsomer concentration. Our investigation here is only preliminary. For example, it would interesting to see how crowding affects the assembly rates at capsomer densities that differ significantly from ρ^*. It may be that the balance between increased effective densities and reduced effective diffusion coefficients changes significantly.

4.2. Modelling $T = 3$ capsids

For a triangulation number of $T = 1$, 60 identical proteins group together in 12 sets of five-fold pentamers. For higher triangulation numbers, the proteins come together in local six-fold coordination as well. Although the chemical makeup of the proteins is identical, it is thought that they may undergo allosteric changes that allow them to have five-fold or six-fold bonding arrangements. Some experiments suggest that capsids can assemble and disassemble from five-fold and six-fold subunits [8, 9]. In that case, it may be a reasonable
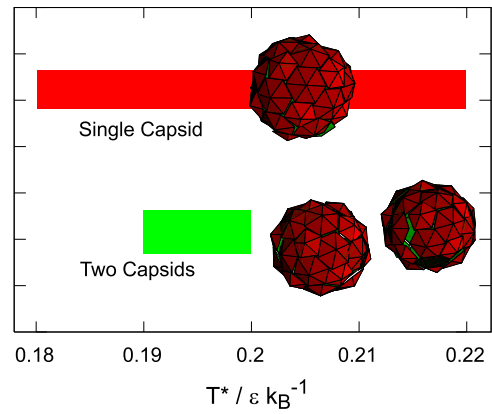
Figure 9. Temperature range for successful assembly of single and double $T = 3$ capsid structures, with $\rho = \rho^*$ and interaction energies described in the text. The coloured bars represent regions where at least one of ten simulations run displayed correct assembly.

approximation to treat the assembly as a hierarchical process, where the pentamers and hexamers first form, and then come together into the capsid.

To model larger triangulation numbers, we introduce hexameric particles with the same side lengths as the pentagonal units, which leads to a radius of $1.18r_b$. The hexamer apex has the same height as the pentagon apex, that is $h = 0.5r_b$, and has the same repulsive interaction. In addition, the interactions between the basal vertices have the same Morse potential form as previously, but with a generalized values of ϵ given by $\epsilon^{55} = \frac{\epsilon}{2}$, $\epsilon^{56} = 2\epsilon$, $\epsilon^{66} = \epsilon$, where ϵ^{ij} is the strength of bonding between a vertex on a capsomer with i sides and a vertex on a capsomer with j sides.

We choose these values to discourage pentamer–pentamer contacts. In [40], Wales and co-workers compared the PES of 20 hexamers and 12 pentamers for a parameterization of their original model to that for the same capsomer population in their newer model, that also includes a selective repulsion site below the plane of the pyramid. For their original model, as well as kinetic traps with inverted capsomers, traps with adjacent pentamers were observed.

We performed similar simulations as before at an effective total capsomer (pentamers and hexamers) density of $\rho = \rho^*$. The simulations were extended to considerably longer times, up to 5×10^8 steps, to allow for more complex assembly mechanisms. Assembly of individual $T = 3$ capsids was observed across a range of temperatures, but, as shown in figure 9, the assembly of two $T = 3$ capsid structures occurred only for a more narrow temperature window. The lower assembly yields for two capsids are due in part to proto-capsid structures being formed without the correct relative number of pentamer and hexamer units. In essence, there is an increased mixing entropy term to contend with. A simulation of just one $T = 3$ capsid automatically sets the right ratio of pentamers to hexamers, but it does not guarantee that two separate proto-capsids will not grow, or that pentamers and hexamers will bind to the right positions.

Kinetic traps were observed, as in the $T = 1$ experiments, with two significant new effects frustrating
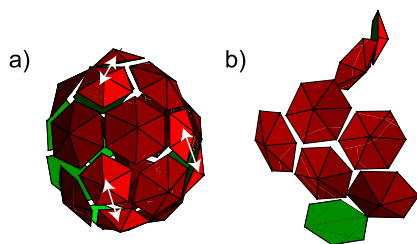
Figure 10. Kinetic traps in the formation of a $T = 3$ capsid. (a) Proto-capsid with adjacent pentamers (shown with arrow pairs), preventing correct geometric formation. (b) A hexameric sheet.
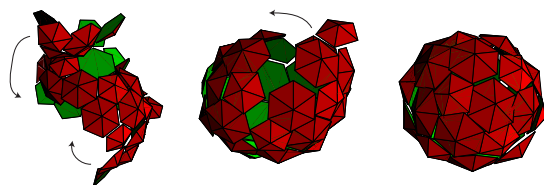
Figure 11. Illustration of the 'shell rearrangement' mechanism described in the text. (left, 8×10^6 cycles) A bonded, misshapen capsid shell has formed. Sections slowly re-orientate to correct geometry (centre, 20×10^6 cycles) A proto-capsid now has one major line defect separating two correctly formed regions. This defect will close, possibly requiring the replacement of some subunits. (right, 400×10^6 cycles) The final structure is formed.

successful assembly in many cases (see figure 10). Firstly, several proto-capsid structures were observed to form with two pentamers adjacent to one another, fixed in place by pentamer–hexamer bonding. This geometric defect then prevented the further formation of the capsid. Secondly, at lower temperatures, hexamers tend to group into planar sheet structures. A similar phenomenon has been observed in experiments [55], where weakening the inter-capsomer bonds of an Adenovirus capsid by treatment with formamide caused the system to initially dissociate into sheet structures.

We also observed more varied mechanisms of assembly for a single $T = 3$ capsid than what was found for the $T = 1$ case of the previous section. For example, we frequently observed large, bonded proto-capsid shells that form quickly in imperfect geometric structures, and then subsequently slowly rearrange to the final icosahedral structure. This is accomplished by the 'closing up' of line defects and repositioning of individual subunits, as illustrated in figure 11.

5. Discussion

We compare computer simulation results for the self-assembly of a model for $T = 1$ virus capsid to PES predictions made recently by Wales [1]. In agreement with the PES picture, we find good assembly yields in regimes where the energy landscape shows a 'funnel'-like topology. Even though the PES calculations were done for a single capsid with all the capsomers connected, the agreement is good, suggesting that this landscape picture may be a fruitful way to analyse how design parameters can aid or hinder self-assembly.

In addition, our simulations reproduce a number of features seen in other simulations of monodisperse self-assembly [19, 23, 25–33, 24], such as a range of temperatures and densities that bound the region of successful assembly, sigmoidal assembly dynamics, hysteresis, and a multitude of kinetic traps including monomer starvation. We also observe that a typical assembly pathway samples many states that are not the lowest energy for that number of particles.

We further extend our model to include crowding agents, and find that these generally lower the yields compared to the optimum at no crowding, but in some cases can increase yields as well. Finally, we extend the model to include hexameric particles, and study the assembly of $T = 3$ capsid structures. There we find that it is more difficult to assemble multiple capsids because there are now two independent species of particles, and mixing entropy terms play a role.

References

[1] Wales D J 2005 *Phil. Trans. R. Soc.* A **363** 357
[2] Fraenkel-Conrat H and Williams R C 1955 *Proc. Natl Acad. Sci. USA* **41** 690
[3] Deng B, OConnor C M, Kedes D H and Zhou Z H 2008 *J. Struct. Biol.* **161** 419
[4] Zlotnick A, Aldrich R, Johnson J M, Ceres P and Young M J 2000 *Virology* **277** 450
[5] Zlotnick A, Johnson J M, Wingfield P W, Stahl S J and Endres D 1999 *Biochemistry* **38** 14644
[6] Suttle C A 2005 *Nature* **437** 356
[7] Caspar D L and Klug A 1962 *Cold Spring Harb. Sym.* **27** 1
[8] Willits D, Zhao X, Olson N, Baker T S, Zlotnick A, Johnson J E, Douglas T and Young M J 2003 *Virology* **306** 280
[9] Hanslip S J, Zaccai N R, Middelberg A P J and Falconer R J 2006 *Biotechnol. Prog.* **22** 554
[10] Fane B A and Prevelige P E Jr 2003 *Adv. Protein Chem.* **64** 259
[11] Carrillo-Tripp M, Shepherd C M, Borelli I A, Venkataraman S, Lander G, Natarajan P, Johnson J E, Brooks C L III and Reddy V S 2009 *Nucleic Acid Res.* **37** D436
[12] Xiao C, Kuznetsov Y G, Sun S, Hafenstein S L, Kostyuchenko V A, Chipman P R, Suzan-Monti M, Raoult D, McPherson A and Rossmann M G 2009 *PLoS Biol.* **7** e92
[13] Casini G L, Graham D, Heine D, Garcea R L and Wu D T 2004 *Virology* **325** 320
[14] Mukherjee S, Thorsteinsson M V, Johnston L B, DePhillips P A and Zlotnick A 2008 *J. Mol. Biol.* **381** 229
[15] Sorger P K, Stockley P G and Harrison S C 1986 *J. Mol. Biol.* **191** 639
[16] Freddolino P L, Arkhipov A S, Larson S B, McPherson A and Schulten K 2006 *Structure* **14** 437
[17] Zlotnick A 1994 *J. Mol. Biol.* **241** 59
[18] Schwartz R, Shor P W and Berger B 2005 *J. Theor. Med.* **6** 81
[19] Nguyen H D, Reddy V S and Brooks C L III 2007 *Nano Lett.* **7** 338
[20] Misra N, Lees D, Zhang T and Schwartz R 2008 *Comput. Math. Methods Med.* **9** 277
[21] Rapaport D C, Johnson J E and Skolnick J 1999 *Comput. Phys. Commun.* **121** 231
[22] Rapaport D C 2004 *Phys. Rev.* E **70** 51905
[23] Rapaport D C 2008 *Phys. Rev. Lett.* **101** 186101
[24] Hagan M F and Chandler D 2006 *Biophys. J.* **91** 42
[25] Elrad O M and Hagan M F 2008 *Nano Lett.* **8** 3850
[26] Chen T and Glotzer S C 2007 *Phys. Rev.* E **75** 51504

[27] Wilber A W, Doye J P K, Louis A A and Lewis A C F 2009 *J. Chem. Phys.* **131** 175102

[28] Nguyen H D, Reddy V S and Brooks C L III 2009 *J. Am. Chem. Soc.* **131** 2606

[29] Wilber A W, Doye J P K, Louis A A, Noya E G, Miller M A and Wong P 2007 *J. Chem. Phys.* **127** 085106

[30] Glotzer S C 2004 *Science* **306** 419

[31] Van Workum K and Douglas J F 2006 *Phys. Rev.* E **73** 31502

[32] Glotzer S C and Solomon M J 2007 *Nat. Mater.* **6** 557

[33] Wilber A W, Doye J P K and Louis A A 2009 *J. Chem. Phys.* **131** 175101

[34] Wales D J 2003 *Energy Landscapes* (Cambridge: Cambridge University Press)

[35] Wales D J and Bogdan T V 2006 *J. Phys. Chem.* B **110** 20765

[36] Leopold P E, Montal M and Onuchic J N 1992 *Proc. Natl Acad. Sci. USA* **89** 8721

[37] Bryngelson J D and Wolynes P G 1989 *J. Phys. Chem.* **93** 6902

[38] Dobson C M 2003 *Nature* **426** 884

[39] Levinthal C 1969 *Mossbauer Spectroscopy in Biological Systems* (Champaign, IL: University of Illinois Press) pp 22–4

[40] Fejer S N, James T R, Hernández-Rojas J and Wales D J 2009 *Phys. Chem. Chem. Phys.* **11** 2098

[41] Tiana G, Sutto L and Broglia R A 2007 *Physica* A **380** 241

[42] Kikuchi K, Yoshida M, Maekawa T and Watanabe H 1991 *Chem. Phys. Lett.* **185** 335

[43] Whitelam S and Geissler P L 2007 *J. Chem. Phys.* **127** 154101

[44] Padding J T and Louis A A 2006 *Phys. Rev.* E **74** 31402

[45] Schwartz R, Shor P W, Prevelige P E and Berger B 1998 *Biophys. J.* **75** 2626

[46] Torrie G M and Valleau J P 1977 *J. Comput. Phys.* **23** 187

[47] Carreira A, Menendez M, Reguera J, Almendral J M and Mateu M G 2004 *J. Biol. Chem.* **279** 6517

[48] Ross P D, Conway J F, Cheng N, Dierkes L, Firek B A, Hendrix R W, Steven A C and Duda R L 2006 *J. Mol. Biol.* **364** 512

[49] Ellis R J and Minton A P 2003 *Nature* **425** 27

[50] Fulton A B 1982 *Cell* **30** 345

[51] del Alamo M, Rivas G and Mateu M G 2005 *J. Virol.* **79** 14271

[52] Stagg L, Zhang S Q, Cheung M S and Wittung-Stafshede P 2007 *Proc. Natl Acad. Sci. USA* **104** 18976

[53] Minton A P 2001 *J. Biol. Chem.* **276** 10577

[54] Ellis R J and Minton A P 2006 *Biol. Chem.* **387** 485

[55] Stasny J T, Neurath A R and Rubin BA 1968 *J. Virol.* **2** 1429