

Torsions

Alex W. Wilber,¹ Jonathan P. K. Doye,^{1,*} and Ard A. Louis^{*2}

¹*Physical and Theoretical Chemistry Laboratory, University of Oxford,
South Parks Road, Oxford OX1 3QZ, United Kingdom*

²*Theoretical Physics, University of Oxford, Keble Road, Oxford OX1 3NP, United Kingdom*

(Dated: December 15, 2008)

This paper is about torsions.

PACS numbers: Don't know

I. INTRODUCTION

The self-assembly of simple building blocks into larger, ordered structures is a ubiquitous process in biology, and holds great promise for applications in materials science and nanotechnology. In general, the dynamic features and end product of a given self-assembly process are entirely determined by the properties of the assembling subunits, and the conditions under which assembly occurs. However, the precise relationships between these determining factors and the consequent assembly processes are incompletely understood. Improving our understanding of the overarching principles of self-assembly would have great value in elucidating a variety of biological processes, as well as offering guidance in the design of synthetic self-assembling systems.

The formation of icosahedral viral protein shells (capsids) provides a particularly inspiring example of self-assembly in nature. A number of these capsid structures can be made to reversibly self-assemble from individual protein subunits (capsomers) *in vitro*, simply by changing the solution conditions. This demonstrates the feasibility of assembling large and complex structures from only a single type of subunit, and in the absence of scaffolding or any other assistive mechanism. Note that although this process is reminiscent of micellar self-assembly it is in fact far more precise, since the completed capsids are monodisperse and highly ordered.

Recently there has been considerable progress in understanding viral capsid assembly, and indeed self-assembly in general. [This paragraph needs expanding, obviously. I've done a lot of reading so it shouldn't take too long.]

Viral capsid subunits are, as proteins (or small groups of proteins), naturally rather structurally and chemically complex. It is interesting to consider which of their properties are important to self-assembly, and how simple a model of a viral subunit could be while still retaining similar assembly properties to those of a real subunit. These questions constitute a major motivation for this work.

Previously, we presented a minimal self assembly model consisting of spherical particles with patchy attractive interactions. The patches were placed so as to favour the formation of icosahedral clusters of twelve particles. Because of the simplicity of the model we were able to run a large number of simulations over a wide re-

gion of parameter space. Our results corresponded with many of the findings of others working on virus assembly. For example, we found that assembly was most efficient when the inter-subunit bonds were relatively weak (or, correspondingly, the temperature was relatively high). Optimal assembly was observed when the specificity of the attractions (modelled in our case as the width of the patches) took a moderate value; high values tended to produce large, relatively disordered clusters, while low values led both to prohibitively long assembly times and relative thermodynamic instability due to a severe entropy penalty. Because of the simplicity of the interactions in the model, other aspects of the results did not correspond to the behaviour of virus capsomers, but are expected to more closely match the behaviour of simple synthetic patchy particles. For example, a novel route to icosahedron formation was found, consisting of the initial formation of large disordered clusters, followed by rearrangement and the “budding off” of completed icosahedra. This is an interesting mechanism, but it is not one which occurs in the formation of viral capsids.

Here we add an additional feature to our previous model, namely the inclusion of torsional constraints (TC) in the potential. Each bond between two particles will have a preferred torsional orientation; rotation of one of the particles around the axis joining the two particles will be disfavoured. This improves the correspondence of our model to the interactions between real capsomers. In general one member of a bound pairs of capsomers is not able to rotate relative to the other because to the complex and rough nature of the surfaces through which they interact. Early synthetic patchy particles, by contrast, are unlikely to be able to implement this constraint. However, we show that the inclusion of TC greatly increases the range of achievable target structures, as well as helping to avoid kinetic traps for simpler structures, and so may represent an important long-term goal for experimentalists.

In this paper we apply our modified model to the assembly of two different shapes, the icosahedron and the dodecahedron. We have already studied the icosahedron in some detail, and so it is a convenient subject for studying the changes caused by using TC. The dodecahedron represents a more difficult assembly target, and we show that TC are essential to its successful assembly. We expect that this is a single example of a trend which will

apply for many if not all larger structures.

II. METHODS

A. Model

The model system we use is a slightly modified version of that used previously⁷. The model consists of spherical particles with a number of sticky patches, which are defined by patch vectors. The interaction potential is pairwise and is based on the Lennard-Jones potential,

$$V_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma_{\text{LJ}}}{r} \right)^{12} - \left(\frac{\sigma_{\text{LJ}}}{r} \right)^6 \right], \quad (1)$$

but the attraction is modulated by a pair of orientationally dependent terms, V_{ang} and V_{tor} . V_{ang} is a function of how directly the two patches are pointing at one another, while the additional factor V_{tor} is a function of the torsional angle between the two particles, i.e. it varies as one of the particles is rotated about the vector connecting the two particles. Thus, the complete potential is

$$V_{ij}(\mathbf{r}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) = \begin{cases} V_{\text{LJ}}(r_{ij}) & r < \sigma_{\text{LJ}} \\ V_{\text{LJ}}(r_{ij}) V_{\text{ang}}(\hat{\mathbf{r}}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) V_{\text{tor}}(\hat{\mathbf{r}}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) & r \geq \sigma_{\text{LJ}}, \end{cases} \quad (2)$$

where $\boldsymbol{\Omega}_i$ is the orientation of particle i . V_{ang} has the form:

$$V_{\text{ang}}(\hat{\mathbf{r}}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) = G_{ij}(\hat{\mathbf{r}}_{ij}, \boldsymbol{\Omega}_i) G_{ji}(\hat{\mathbf{r}}_{ji}, \boldsymbol{\Omega}_j) \quad (3)$$

$$G_{ij}(\hat{\mathbf{r}}_{ij}, \boldsymbol{\Omega}_i) = \exp \left(-\frac{\theta_{k_{\min}ij}^2}{2\sigma_{\text{ang}}^2} \right), \quad (4)$$

where σ_{ang} gives the width of the Gaussian, θ_{kij} is the angle between patch vector k on particle i and the interparticle vector \mathbf{r}_{ij} , and k_{\min} is the patch that minimizes the magnitude of this angle. Hence, only the patches on each particle that are closest to the interparticle axis interact with each other.

In the calculation of V_{tor} , an additional reference vector on each particle needs to be defined. In our simulations, we have chosen this reference vector in each case to lie on the symmetry axis of the particle and to point away from the mean direction of the patch vectors. V_{tor} is a maximum when the projections of the two reference vectors onto the plane perpendicular to the interparticle vector lie parallel. Specifically, if θ_{tor} is the angle between the projections,

$$V_{\text{tor}} = \exp \left(-\frac{\theta_{\text{tor}}^2}{2\sigma_{\text{tor}}^2} \right) \quad (5)$$

where \mathbf{v}_i and \mathbf{v}_j are the reference vectors on particles i and j respectively, and σ_{tor} gives the width of the Gaussian. The effect of the inclusion of V_{tor} in the potential is

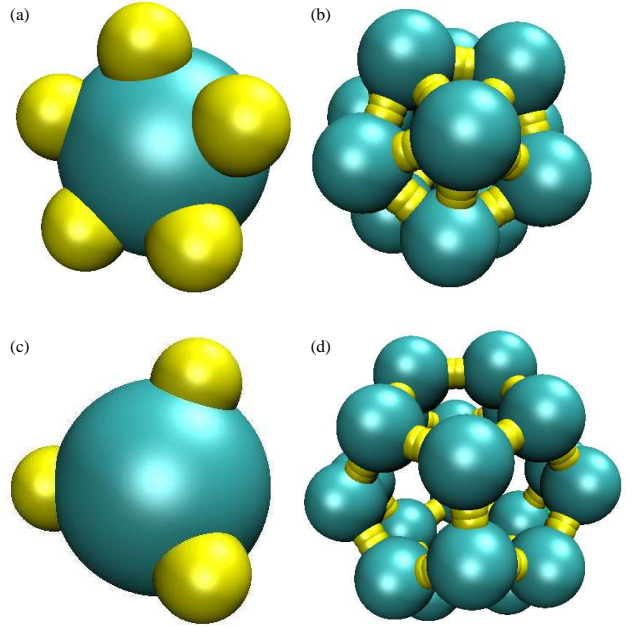


FIG. 1: (Colour Online) Monomers designed to assemble into each of our two target structures, along with a complete example of each. (a) A particle designed to form part of an icosahedron, (b) a complete icosahedron, (c) a particle designed to form part of a dodecahedron, and (d) a complete dodecahedron. The yellow spheres indicate the orientations of the patch vectors, and are not particles in themselves.

to penalise twisting around the interparticle vector, with smaller values of σ_{tor} giving a stronger constraint.

Both the positions of the patch vectors and that of the reference vector are placed in each case so that the target structures represent an energy minimum. The patch vectors are chosen such that they point directly at the neighbouring particles in the target structure, while the positioning of the reference vector described above leads to structures of high symmetry having the lowest energy, as desired. The two target structures, icosahedra and dodecahedra, are shown in Fig. 1, with the positions of the patch vectors and the reference vector marked. [This last bit isn't true, but is a good idea perhaps.]

B. Monte Carlo simulation

Our simulations make use of the Metropolis Monte Carlo (MC) algorithm in the canonical ensemble, with fixed numbers of particles, volume and temperature. In each MC move, a single particle is chosen and either a translation or a rotation move is attempted. A translation move translates a particle to a point chosen from a cube centred on the particle's previous position. The rotation moves make use of a quaternion representation for particle orientation, adding a small random perturbation to the particle's current quaternion before renormalising

it.

This algorithm does not directly simulate physical trajectories, and is more often used to obtain thermodynamic averages for systems in equilibrium. However... [I want to say *exactly* the same thing as in the last paper - should I just copy and paste, or try to write the same thing using different words, or perhaps even refer back to the last one?]

[If I'm going to include the dodecahedron landscapes, need to explain parallel torsioning as well.]

C. Umbrella sampling and parallel torsioning

[I don't think we can call it parallel torsioning.]

D. Cluster determination and ring counting

[Guess I will need to explain the ring counting, at least briefly, if I'm going to include the landscapes.]

III. RESULTS

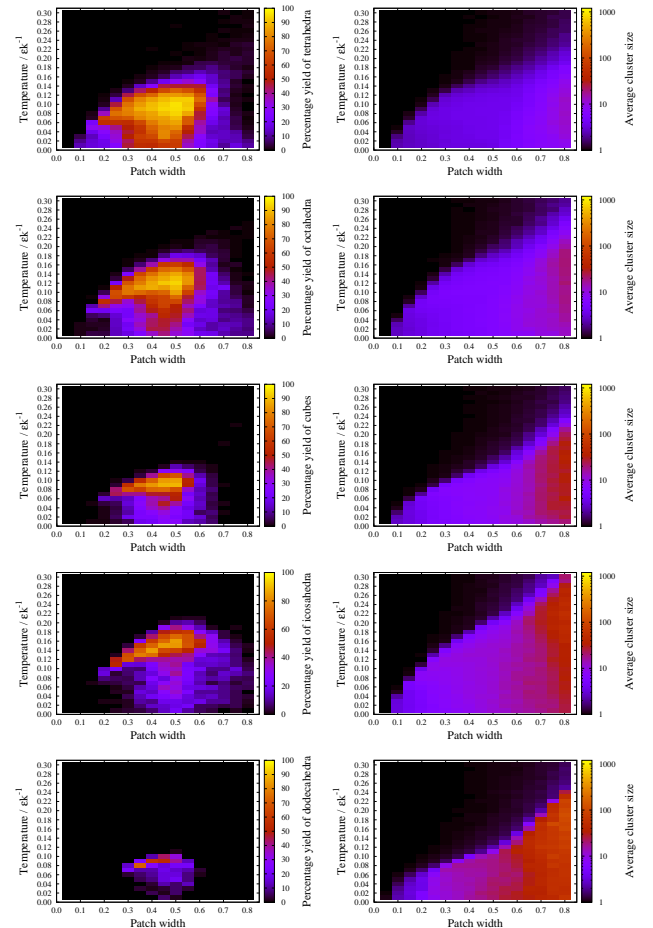
A. Energetics

[Are we going to use gmin plots? It might be interesting to compare high patch-width plots for TC/no-TC simulations, since this will succinctly show the fact that TC cause icosahedra to remain stable to much higher σ_{ang} .]

B. Self-assembly of icosahedra

In order to obtain a clear picture of the effect of including TC (torsional constraints) in our model we carried out two sets of simulations over a range of temperatures and patch widths σ_{ang} using particles designed to assemble into icosahedra. In one set of simulations, the torsional flexibility σ_{tor} was set to infinity (i.e. no TC were present), while in the other set we chose $2\sigma_{\text{tor}} = \sigma_{\text{ang}}$, so that the angular and torsional flexibilities were coupled. [Want to explain the factor of two, but feel that saying it's arbitrary won't look too hot.] Results for both the yields of icosahedra and the average sizes of clusters formed are shown in Fig. ??.

Those results where $\sigma_{\text{tor}} = \infty$ have been presented before[?], and are included here for comparative purposes. We shall briefly outline the general features of the results for the $\sigma_{\text{tor}} = \infty$ case, before moving onto the new results for finite σ_{tor} . It can be seen that there is a region of parameter space in which assembly of icosahedra clusters is successful, which is bordered by three different regions in which assembly is unsuccessful for different reasons. At high temperatures, the entropic cost of forming clusters becomes prohibitive, and the stable state is one in



which all particles exist either as monomers or in small clusters. At very low temperatures icosahedra are not formed because of kinetic trapping; instead, the particles form extended networks. These networks are neither entropically nor energetically favourable, but the temperature is too low for bonds to be broken in the timescale of the simulation, and so the system is not able to rearrange to a more favourable state. Finally, at high patch widths σ_{ang} , icosahedral clusters cease to be the most stable state. The particles effectively have attractive interactions over their entire surface, and so form large, liquid-like droplets in order to maximise the number of nearby particles with which they can interact.

The introduction of TC dramatically changes the behaviour of the system. This change is particularly visible in Fig. ??(d). The system no longer forms any large clusters, either at low temperature or at high patch width. In general, icosahedron-forming particles will tend to form convex shapes because of the placement of their patches. However, in the absence of TC some particles may be positioned “upside-down”, such that they have patches pointing outwards from their cluster and permitting the growth of extended structures. When TC are applied this is no longer possible, so that only convex structures are formed and large clusters are almost entirely absent. We

see some larger clusters forming at high values of σ_{tor} , where the TC are largely relaxed, but otherwise cluster sizes above twelve are rarely seen.

The inhibition of large clusters has a number of effects on the formation of icosahedra. Some of the main classes of structure competing with the icosahedra are no longer viable. At low temperatures, particles cannot now form extended networks, and hence remain mobile and able to contribute to the assembly of icosahedra. The situation is similar at moderately high values of σ_{ang} , where the formation of liquid-like droplets is no longer favourable. These effects result in increased yields of icosahedra both at low temperatures and at high patch widths.

The introduction of TC also has some detrimental effects on assembly. In the absence of TC, assembly of icosahedra proceeds by two mechanisms. In the first mechanism, the icosahedra are built up by addition of monomers and small clusters, until complete icosahedra are formed. In the second, the particles first form large aggregates, which then rearrange and “bud off” completed icosahedra. The first mechanism dominates at temperatures close to T_{clust} , the temperature above which the stable state is a gas of monomers, while the second mechanism is more important at lower values of T , and at higher patch widths. The optimal conditions for assembly are those in which both mechanisms contribute. However, when TC are introduced the formation of aggregates is prevented, and so the contribution of the budding mechanism is eliminated. This explains the very different shapes of the regions of successful assembly with and without TC.

Despite the prevention of the formation of large networks, assembly at low temperature in those simulations with TC is still generally poor. Instead of forming networks, the particles rapidly form clusters which begin to grow towards forming icosahedra. However, this leads to a case of “monomer starvation”, where there are few monomers left to enable continued cluster growth. The result is a large number of partially formed clusters which are unable to form complete icosahedra. Because of the low temperature they are unlikely to break apart, and so the yield remains low over the timescale of the simulation.

[Still to explain: Steeper slope at low σ . Higher max. temp..]

C. Mechanisms of assembly

Fig. 2 shows the average cluster size as a function of simulation time for simulations both with and without TC at a range of temperatures and at $\sigma_{\text{ang}} = 0.45$. This plot highlights one of the key changes in the assembly process caused by the addition of TC. In the absence of TC, over a broad range of temperatures the average cluster size passes through a maximum, before decreasing back down towards an equilibrium value of around twelve. The system initially forms large clusters, which

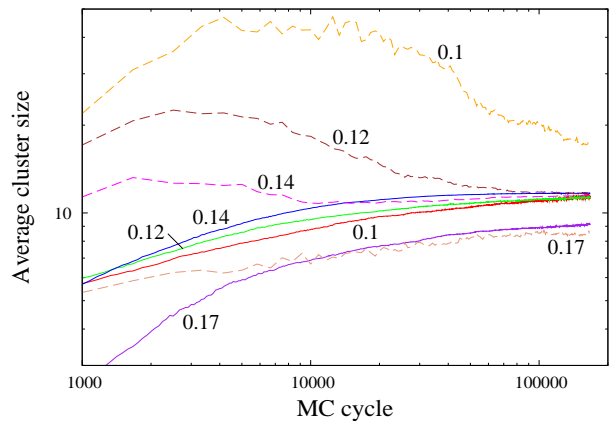


FIG. 2: (Colour Online) The mean cluster size (sampled over particles) as a function of number of MC cycles at $\sigma_{\text{ang}} = 0.45$ and at different temperatures (as labeled) for a system of 1200 particles at a number density of $0.15\sigma_{\text{LJ}}^{-3}$. The dotted lines are for simulations without TC, while the solid lines are for simulations with TC, $\sigma_{\text{tor}} = 0.111$, i.e. $2\sigma_{\text{tor}} = \sigma_{\text{ang}}$. Each line is an average over ten simulations.

then rearrange and break up into completed icosahedra. At low temperatures ($T \leq 0.1$) this process of rearrangement can be extremely slow. In the presence of TC, however, the formation of large clusters is strongly inhibited. Instead, assembly proceeds only by the addition of monomers and small clusters, so that the average cluster size increases monotonically towards 12.

At the relatively high temperature of $T = 0.17$ assembly becomes nucleation-limited both for simulations with and without TC, since it depends on the occasional formation of clusters large enough not to quickly decompose back to monomers. Once formed these clusters tend towards completion, leading to a gradual increase in the mean cluster size. In this case larger clusters are generally not formed, regardless of the inclusion of TC, because the formation of larger clusters is dependent on the formation of strained bonds which will be highly unstable at high temperature.

Another perspective on the dynamics of the assembly process is given by Fig. 3, which shows the populations of particles in clusters of different sizes as a function of time. [Unfortunately I don’t think this is quite worth writing yet, since the plot will be much more interesting once we have data from a *low* temperature run, where we can hopefully see clear evidence of monomer starvation. I’ll then use this to refer again to how assembly at low T is prevented by monomer starvation rather than gelation.]

D. Thermodynamics

In order to obtain thermodynamic data on our model systems, we made use of the umbrella sampling technique described in Sec. II C. The resultant plots of heat capac-

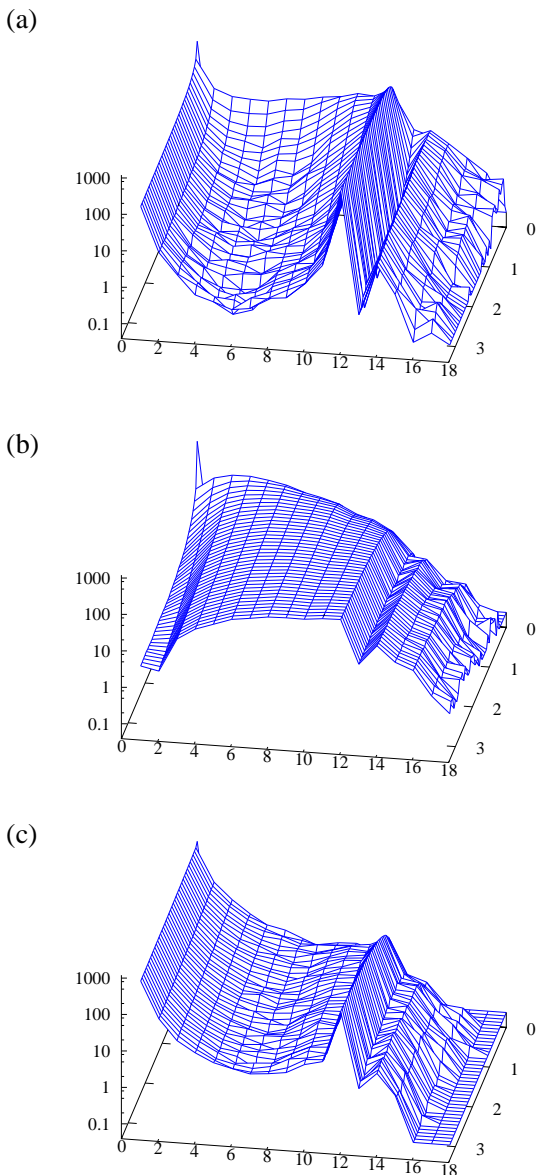


FIG. 3: (Colour Online) The populations of particles in clusters of different sizes n as a function of number of MC cycles under different conditions, averaged over 100 repetitions. A log scale is used so that the very small populations of intermediates are discernible. States with zero population have been set to a minimum value of 0.1 to aid in visualisation. Each simulation consists of 1200 particles at a number density of $0.15\sigma_{\text{LJ}}^{-3}$. The temperatures and patch widths used are (a) $T = 0.15$, $\sigma_{\text{ang}} = 0.45$, the optimal conditions for assembly, and (b) $T = 0.17$, $\sigma_{\text{ang}} = 0.4$, close to the transition to a monomer gas. In each case TC were included, with $2\sigma_{\text{tor}} = \sigma_{\text{ang}}$. [Due to monstrous technical difficulties, the axis labels aren't yet included. They are, from left to right, the population P_n of particles in clusters of size n , the cluster size n , and the MC cycle / 10^4 . Also I would like to include a simulation at lower temperature, where kinetic trapping is a problem, and possibly one at optimal conditions without torsions for comparison.]

ity as a function of temperature are shown in Fig. ??[not made plot yet] (a).

The same data also allows plotting of the equilibrium yield of target structures as a function of temperature. This data is shown in Fig. [not done yet] (b) along with the actual yields obtained in dynamic simulations.

It is interesting to consider how favourable it is for a particle to form part of a cluster of a certain size, as opposed to remaining free. In general, particles in clusters will tend to have a lower entropy but also a lower energy than free particles, and these quantities will also vary between different sizes of cluster. It is possible to define a free energy to quantify the preference of particles to occupy clusters of particular sizes. If one considers one particular particle in the system to be the “reference particle”, the states of the system can be divided according to the size i of the cluster occupied by that particle, and each of these sets of states can then be assigned a free energy based on how frequently they occur, i.e.

$$A_i = -\beta \ln p_i + c \quad (6)$$

where A_i is the free energy of the set of states in which the reference particle occupies a cluster of size i , and p_i is the probability of the reference particle being in such a cluster. The free energies may be found only to within an additive constant c which is the same for all clusters. The quality of sampling can be increased for a single simulation run by considering each particle in turn to have been the reference particle. This results in a simple equation:

$$A_i = -\beta \ln \left(\sum_{t=1}^{N_t} n_{i,t} \right) \quad (7)$$

where the sum is over all the timesteps in the simulation, and $n_{i,t}$ is the number of particles in a cluster of size i at timestep t .

These free energy plots are shown in Fig. ??[also not done yet - a little tricky].

E. Assembly of dodecahedra

Dodecahedra are largest of the Platonic solids, and for a number of reasons represent a far more difficult assembly target than icosahedra. Heat maps showing the yields and average cluster sizes for simulations of dodecahedron-forming particles with and without TC are shown in Fig. 4. In the absence of TC, successful assembly of dodecahedra is never observed. We cover the reasons for this in detail elsewhere⁷, but the assembly is essentially prevented by strong competition from misformed clusters. Misformed clusters become more important competitors of larger structures because the set of possible misformed structures grows exponentially with the number of particles, and consequently the set as a whole becomes more entropically stable. The low total binding energy of

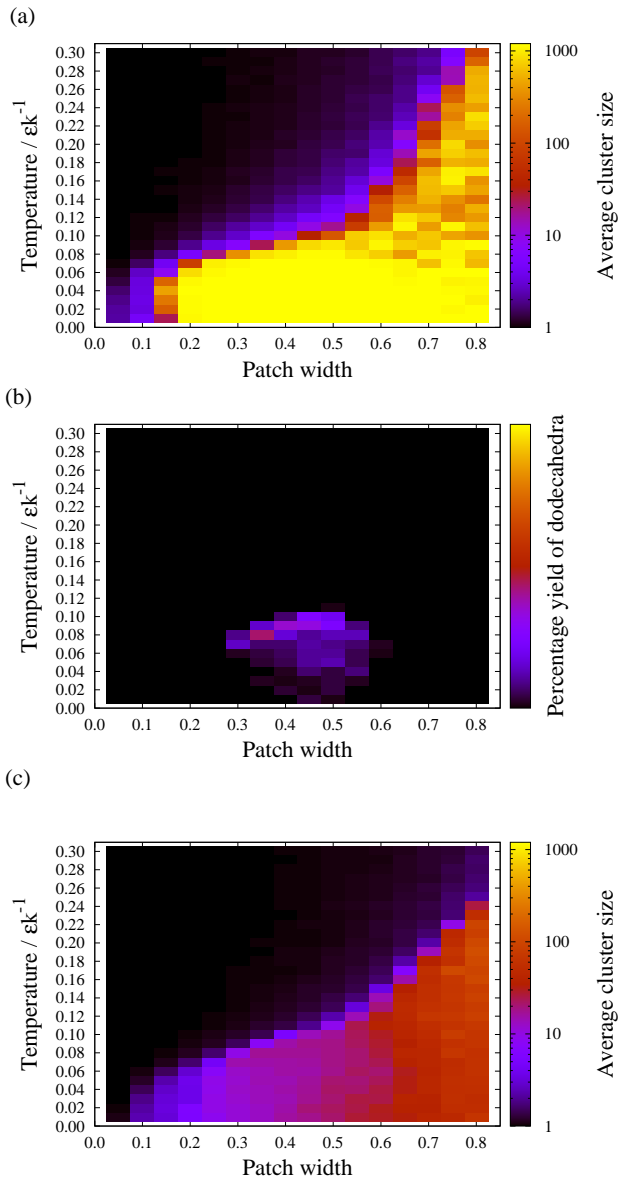


FIG. 4: (Colour Online)

the dodecahedral cluster (having only three patches per particle) makes it particularly vulnerable to competition from other states.

Once TC are introduced, most of the possible erroneous structures become strongly disfavoured, and successful assembly of dodecahedra takes place over a limited range of parameters. Under optimal conditions, the final yield of dodecahedra after 80 000 MC cycles is [X]%, and at long times the yield rises to [Y]%. In the presence of TC, it is in fact difficult to construct any erroneous structures at all, and so the main competition arises at low temperatures from monomer starvation leading to the formation of incomplete structures, while at high temperatures a monomer gas becomes the stable state. At

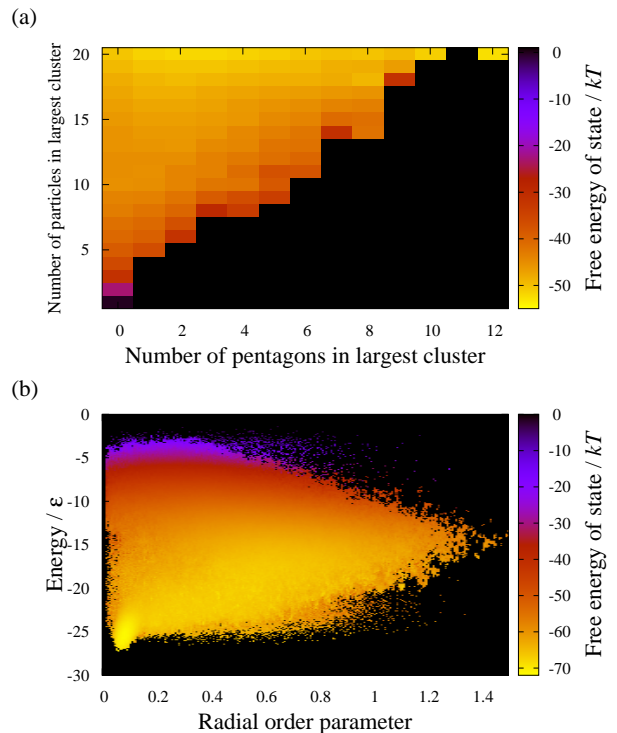


FIG. 5: (Colour Online) [These plots, produced at the transition temperature where dodecahedra become stable in each case, are **obviously nonsense** - I think there must be something wrong with the script that produces them. Now is not the moment for debugging though, and hence these are just rather pretty placeholders. The first is supposedly without TC, and the second with.]

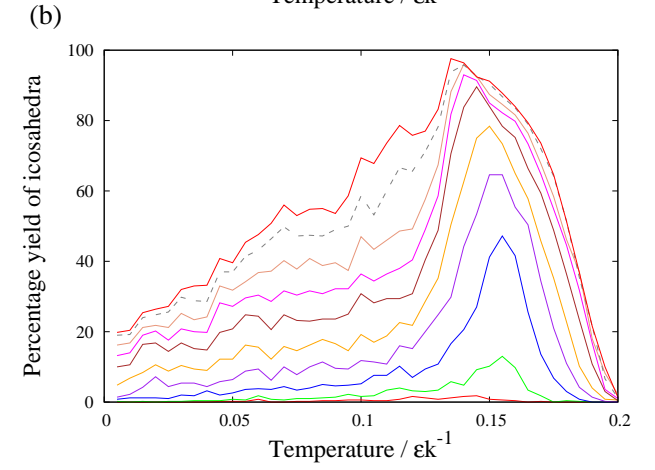
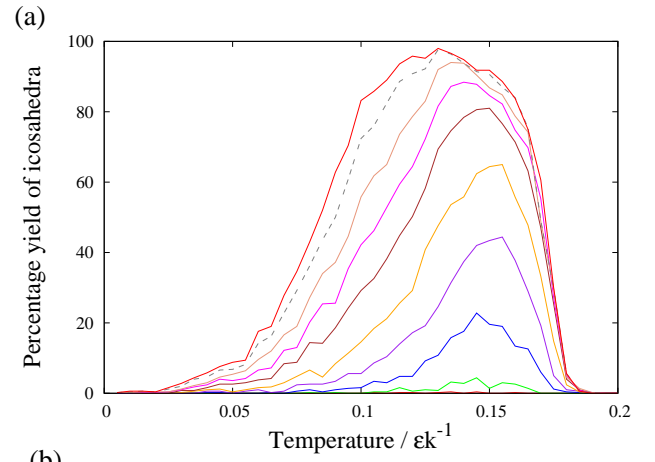
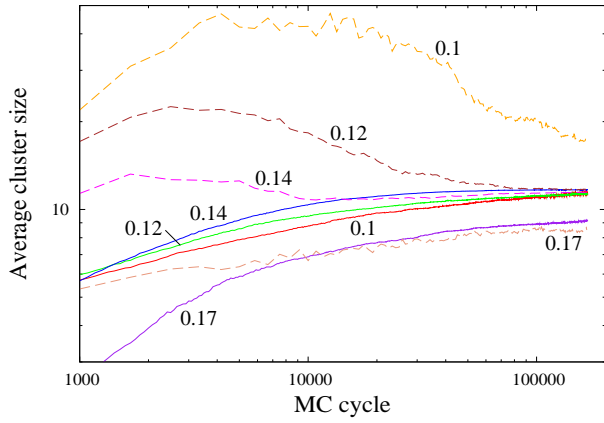
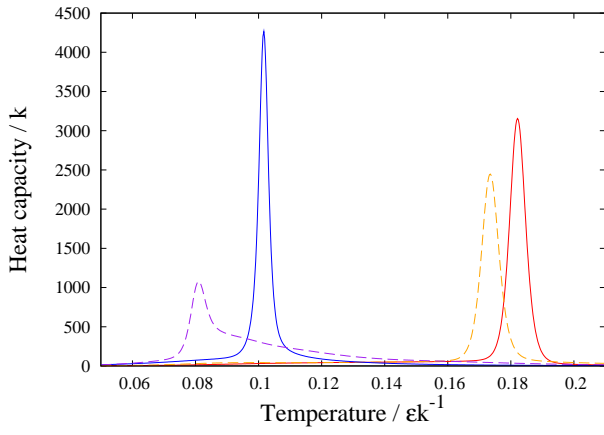
high values of σ_{ang} and σ_{tor} , of course, both the angular and TC begin to break down, and so a variety of malformed structures are still observed.

By use of the methods of umbrella sampling and parallel torsioning, described in Section II C, we were able to obtain equilibrium statistics for a system of 20 dodecahedron-forming particles (sufficient for a single dodecahedron). Fig 5 shows a pair of energy landscapes, for systems with and without TC. [I anticipate having much more to say on this once decent plots are done.]

Our findings with dodecahedra show that TC are very important in the assembly of large structures. Indeed, we expect that torsional requirements will be a prerequisite for the successful assembly of any large or complex structures, and that icosahedra are probably the largest clusters that can be reliably formed without.

IV. CONCLUSIONS

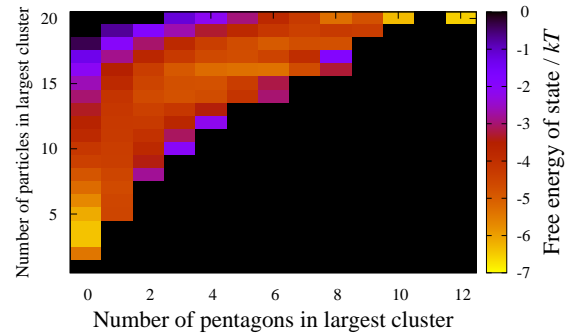
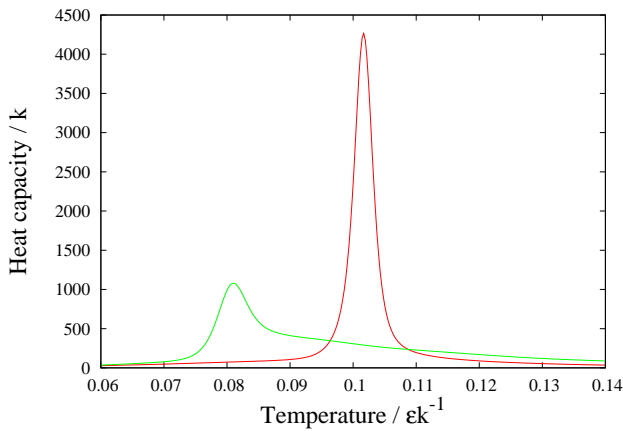
- Inhibits formation of large aggregates (which is both beneficial and detrimental, since it prevents budding) - Greatly reduces set of erroneous structures that can be



formed - More accurately matches real-life assembly of viruses - Probably hard to implement in synthetic systems (pair of patches), but ultimately important, because
 - Allows formation of larger/more complex structures]

Acknowledgments

The authors are grateful for financial support from the EPSRC and the Royal Society.



Are we still grateful to the Royal Society?

* Authors for correspondence