

ARTICLE

Designing Hierarchical Molecular Complexity: Icosahedra of Addressable Icosahedra

Szilard N. Fejer^a, Rosemary G. Mantell^b, and David J. Wales^b

^a Provitam Foundation, 16 Caisului Street, Cluj-Napoca, Romania;

^b University Chemical Laboratories, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.;

E-mail: dw34@cam.ac.uk

ARTICLE HISTORY

Compiled January 20, 2018

ABSTRACT

Clusters and shells based upon icosahedral symmetry are characterised, constructed from component particles that retain memory of their neighbours in a specific reference structure. This memory provides the particles with ‘addressable’ characteristics, with a ground state corresponding to the structure where all the components are correctly addressed in terms of the local environment. The interparticle potentials have separate attractive and repulsive components, defined by the reference structure. For a single target structure, the relaxation efficiency is mostly determined by the variation of the attractive component for correctly and incorrectly addressed particles. The effects of varying the addressability can be visualised directly in terms of the underlying energy landscape, and follow quantitative predictions from catastrophe theory. A doubly-addressable landscape can be designed within the same framework. In well-defined regions of parameter space the predicted global minima for aggregates of the target cluster (target monomer) form ‘superclusters’, which can be described in terms of multiple interacting copies of local minima for the monomer. The predicted lowest energy superclusters formed from aggregates of addressable icosahedral clusters and shells are themselves based on icosahedral packing. These hierarchical icosahedral structures should be realised experimentally if particles can be synthesised to match the interactions encoded in the addressable potentials.

KEYWORDS

hierarchical self-assembly; addressable clusters

1. Introduction

As for familiar macroscopic length scales, creating a molecular machine requires assembly of components into specific relative positions. Such components are termed ‘addressable’, since the target design is encoded in their interactions via a reference structure, and hence in the underlying potential energy landscape. One promising route for addressable self-assembly employs DNA-coated nanoparticles, in which the choice of DNA sequences determines the aggregation [1], and complex phase behaviour can be designed with such particles [2]. Self-assembly of thousands of distinct ‘DNA bricks’ into a target

structure has been achieved experimentally [3], initiating an active new field of research. Theory and simulation can play an important role here, in elucidating design principles and helping to define essential features that would endow particles with the capability to serve as addressable components. Models that reproduce key elements of the DNA brick experiments have already been presented [4], and annealing schedules that optimise the yield of a target aggregate have been predicted [5]. Colloids may offer an alternative route to realisation of new addressable materials, and here again, experimental advances [6, 7] have stimulated a variety of theoretical and computational studies [8–12].

Ensuring the kinetic accessibility of target addressable structures is still a challenging task for both experiments and theory [13]. To facilitate the design of addressable colloids, it is important to understand the features of the interparticle potential that affect the thermodynamic and kinetic efficiency of assembly into a target structure. The yield of a target aggregate can be increased using non-stoichiometric particle concentrations, with ratios selected to favour the ideal aggregation pathway [14]. This approach is especially important in addressable assembly with strong short-range interactions (e.g. ‘sticky’ colloids or irreversible aggregation). However, if the effective interparticle interactions are longer-range, the self-assembling features of the underlying energy landscape become increasingly important.

Hierarchical self-assembly from addressable components adds another level of complexity. There are at least three types of addressable systems capable of two-level hierarchical self-assembly: (1) The first level target structures are obtained through self-assembly, and act as an addressable components for the second level assembly; (2) the first level targets are composed of addressable particles, which further self-assemble into clusters or crystals; (3) the first level targets are addressable, and they also correspond to addressable monomers blocks for the next level of self-assembly. The conditions required to enhance self-assembly for case (1) have recently been investigated [15]. Here we focus on case (2), looking at the minimal modifications required for an isotropic pair potential to make it addressable and support hierarchical self-assembly. We also investigate how the global energy landscape is affected if we systematically decrease the attraction between particles that are not nearest neighbours in the target structure.

A simple addressable potential, where the magnitude of non-nearest-neighbour interactions is reduced according to the environment in the target reference structure, was recently shown to provide a theoretical route to addressable clusters [12]. In this formulation the particles interact via pair potentials, and addressability was defined by having different interaction strengths between pairs of particles that are nearest-neighbours in the reference structure, relative to pairs that are non-nearest-neighbours. This memory of the local environment in the reference structure was achieved by reducing the magnitude of the attractive part of the potential between non-nearest-neighbour particles.

Here we introduce a more flexible formulation, which enables us to design greater specificity into the potential energy landscape of the target structure, which we will refer to as the target monomer when considering multiple copies of the target in an aggregate. To summarise the terminology: the

components of the monomer will be referred to as particles, which might be individual colloids or DNA bricks. Aggregates of monomer structures will also be referred to as superclusters. The monomers within a supercluster need not have the correct monomer target structure, but for suitable regions of parameter space, the predicted global minima for aggregates do indeed consist of weakly perturbed copies of the target, with all the component particles addressed correctly relative to the reference.

We find that the favoured morphologies of aggregate superclusters can themselves be highly ordered for well defined regions of parameter space. To illustrate the simplicity and generality of this scheme we show how it can be applied to target monomers composed of particles bound by isotropic pair potentials, and particles based on discoids. The two reference structures for the target monomer are then based either on a 13-atom Mackay icosahedron [16] or a 12-particle shell, where the component discoids could represent pentameric capsomers in a capsid with triangulation number $T = 1$ [17]. The effect of varying parameters to optimise the self-assembling characteristics of these targets can be visualised using disconnectivity graphs [18, 19]. As the potential energy landscape becomes more funnelled towards the target it also becomes smoother, as higher-lying local minima merge with transition states via fold catastrophes [20]. We show that the landscape can be simplified in the same way for a doubly-addressable potential, which targets two specific permutational isomers corresponding to competing close-packed and icosahedral morphologies in a double-funnel system.

For suitable parameter choices we find that aggregates of the two types of icosahedral targets are composed of well-defined target monomers, having each addressable particle correctly sited within each monomer. To achieve this organisation with 13-atom icosahedral clusters, the addressable potential requires terms that tend to keep apart atoms with the same address. For the 12-particle icosahedral shells, the discoids already have a polarity, which distinguishes between the inside and outside surfaces of the shell. This Janus character helps to retain the integrity of the target clusters when they aggregate together. The likely global minima for these aggregates are characterised systematically as a function of size up to superclusters of 13 target monomers. For larger aggregates we investigate sizes where higher order icosahedral organisation is possible, and find that global optimisation does indeed lead to this motif for clusters of 55 and 147 monomers. Icosahedra of icosahedra are also probably the global minima for aggregates of 309 and 561 monomers, containing 4017 and 7293 atoms in total, respectively.

2. Methods

2.1. *Interparticle Potentials*

Atomic Clusters. We first consider addressable potentials based on the pairwise Lennard-Jones (LJ)

function [21], with separate coefficients for the repulsive and attractive terms:

$$V = 4 \sum_{i < j} \left[\epsilon_{ij}^{\text{rep}} \left(\frac{\sigma}{r_{ij}} \right)^{12} - \epsilon_{ij}^{\text{att}} \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (1)$$

where r_{ij} is the distance between atoms i and j . Each atom is then programmed with knowledge of the local environment for a reference structure in which the distances are r_{ij}^0 :

$$\epsilon_{ij}^{\text{rep}} = \begin{cases} \epsilon_{\text{NN}}^{\text{rep}}, & r_{ij}^0 \leq r_c, \\ \epsilon_{\text{NNN}}^{\text{rep}}, & r_{ij}^0 > r_c, \end{cases} \quad \epsilon_{ij}^{\text{att}} = \begin{cases} \epsilon_{\text{NN}}^{\text{att}}, & r_{ij}^0 \leq r_c, \\ \epsilon_{\text{NNN}}^{\text{att}}, & r_{ij}^0 > r_c. \end{cases} \quad (2)$$

The cutoff, $r_c = 1.2$, distinguishes nearest-neighbour contacts (NN) and non-nearest-neighbours (NNN). To design a potential with multiple addresses we use the minimum value of r_{ij}^R over all the reference structures, R , in place of r_{ij}^0 . This approach has some features in common with Gō models [22–24] and potentials that employ overlap with a reference configuration [25–37]. Our focus in the present contribution is on addressable building blocks, where we are predicting rather than designing the higher-order structures that will be formed on aggregation of these building blocks. This approach is complementary to schemes that seek design principles for self-assembly of a target higher-order configuration. For such targets, inverse statistical-mechanical methods have been employed to optimise the form of an isotropic pair potential to produce crystals with low coordination number [38], including diamond and wurtzite lattices [39, 40], as well as square and three-coordinate graphene structures [41–43]. These designs are remarkable because they only involve isotropic interactions, without specific directional characteristics, which may facilitate experimental realisations for colloids [38]. The higher-order structures that we predict in the present work for aggregates of addressable monomers are an emergent property rather than a design. The resulting hierarchical icosahedral morphologies (superclusters) are composed of identifiable addressable monomers with relatively high coordination numbers, in contrast to the low coordination numbers designed using optimised pair potentials.

Discoidal Particles. The second potential we consider corresponds to discoidal building blocks. We have previously used the model introduced by Paramonov and Yaliraki [44] (PY) to design a variety of self-assembling structures, including shells, tubes and helices [45, 46]. A frustrated bilayer design successfully reproduced the twisted and spiral ribbons and kinked morphologies observed for decorated amyloid fibres and Bauhinia seedpods, with the same trends observed for helices constructed from ellipsoidal magnets [47]. Left-handed building blocks composed of symmetrical pairs of oblate discoids can be designed to assemble into both left- and right-handed helices [48]. Hence the PY potential provides an outstandingly versatile framework for mesoscopic modelling.

The PY formulation can be viewed as an ellipsoidal generalisation of the LJ potential, with separate

shape parameters defining the attractive and repulsive components, and a distance term defined using the elliptic contact function [44]. The functional form is significantly more complicated than the simple isotropic LJ function (SI), but addressability can be defined in the same way, using separate scale factors for the attractive and repulsive contributions corresponding to NN and NNN contacts. Details of the ellipsoidal parameterisation and the two axial sites that define these building blocks are given in the SI.

Aggregates. For multiple copies of a particular monomer containing N particles we simply repeat the addressable potential for each copy, by treating the particle index modulo N in the repulsive and attractive parameters (SI). New parameters $\epsilon_{\text{self}}^{\text{rep}}$ and $\epsilon_{\text{self}}^{\text{att}}$ are introduced for particles that share the same address.

Multiple copies of the target cluster were considered for 13 LJ atoms and 12 PY ellipsoids, setting the reference structures to the global minimum Mackay icosahedron [16] and the icosahedral $T = 1$ shell [17], respectively. These addressable clusters are denoted $\text{LJ}_{13}^{I_h}$ and $\text{PY}_{12}^{I_h}$, and the aggregates containing M copies of each particular target monomer are denoted $[\text{LJ}_{13}^{I_h}]_M$ and $[\text{PY}_{12}^{I_h}]_M$. The repulsive term that acts between particles with the same address is needed to obtain aggregates of the addressable $\text{LJ}_{13}^{I_h}$ cluster where the target monomers can be distinguished. For $\epsilon_{\text{self}}^{\text{att}} = 0$ a value of $\epsilon_{\text{self}}^{\text{rep}} = 10^4$ produces a clear separation of target monomers for $\text{LJ}_{13}^{I_h}$. The large value arises because the repulsion needs to have a significant effect at relatively large separations. Alternative formulations are possible using longer-ranged terms. The discoidal building blocks used to construct the $\text{PY}_{12}^{I_h}$ shell already have a two-faced Janus character, to distinguish the inside and outside of the shell [46]. For $[\text{PY}_{12}^{I_h}]_M$ aggregates a value of $\epsilon_{\text{self}}^{\text{rep}} = 1$ proved sufficient to produce assemblies composed of distinct target monomers. Hierarchical aggregation depends on the balance of energy and length scales for interactions within and between the monomers, all governed by the pair interactions between the addressable particles that self-assemble into the target monomers. The $\text{LJ}_{13}^{I_h}$ and $\text{PY}_{12}^{I_h}$ target monomers contain 13 and 12 distinct addressable particles, respectively.

2.2. Exploring and visualising the energy landscapes

The energy landscapes for single and multiple copies of the target monomers were explored using established techniques based on geometry optimisation within the computational potential energy landscapes framework [49]. Structure prediction employed basin-hopping global optimisation [50, 51], starting from a set of random configurations and from candidates obtained by explicit construction. Kinetic transition networks were sampled by systematic searches for transition states and pathways connecting local minima using the doubly-nudged [52] elastic band [53, 54] approach and hybrid eigenvector-following [55, 56]. These networks enable us to visualise the potential or free energy landscape using disconnectivity graphs [18, 19]. The vertical axis corresponds to potential energy in these diagrams, and each branch terminates at the energy of a local minimum. The minima are connected when they can

interconvert via a pathway where the highest transition state lies below a given energy threshold, and these connections are performed at a regular series of thresholds chosen to provide insight into the overall structure. The branches are positioned on the horizontal axis to avoid overlap, which groups minima in a hierarchical fashion. The organisation of the landscape then enables us to predict whether the system will be able to relax efficiently to the global minimum [19, 49] from any starting configuration. Our aim is to design building blocks with encoded interactions that promote efficient self-assembly, without any external input. Examples are illustrated in Figures 1-3.

3. Results and Discussion

3.1. Assembly of icosahedral target monomers

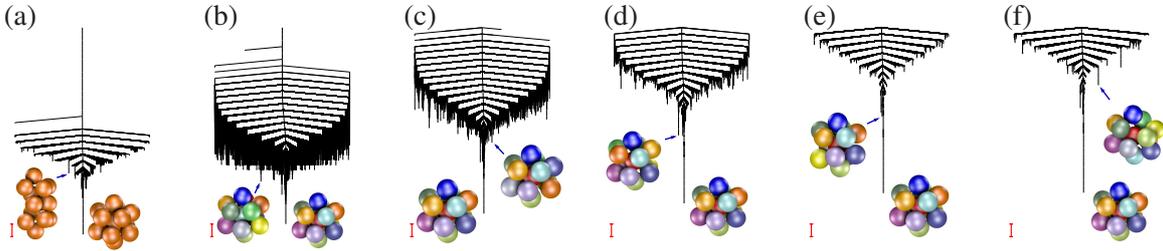


Figure 1. Disconnectivity graphs for addressable $LJ_{13}^{I_h}$ clusters. $\epsilon_{NN}^{\text{att}} = \epsilon_{NN}^{\text{rep}} = \epsilon_{NNN}^{\text{rep}} = 1$ in each case, except for (f) where $\epsilon_{NN}^{\text{rep}} = 1.5$. (a) $\epsilon_{NNN}^{\text{att}} = 1.00$, (b) $\epsilon_{NNN}^{\text{att}} = 0.75$, (c) $\epsilon_{NNN}^{\text{att}} = 0.50$, (d) $\epsilon_{NNN}^{\text{att}} = 0.25$, (e) $\epsilon_{NNN}^{\text{att}} = 0.00$, (f) $\epsilon_{NNN}^{\text{att}} = 0.00$. The scale bar corresponds to one unit of potential energy and is the same height for each panel. All permutation-inversion isomers are lumped together in panel (a), which corresponds to the usual Lennard-Jones potential. In the other panels enantiomers are paired in each branch.

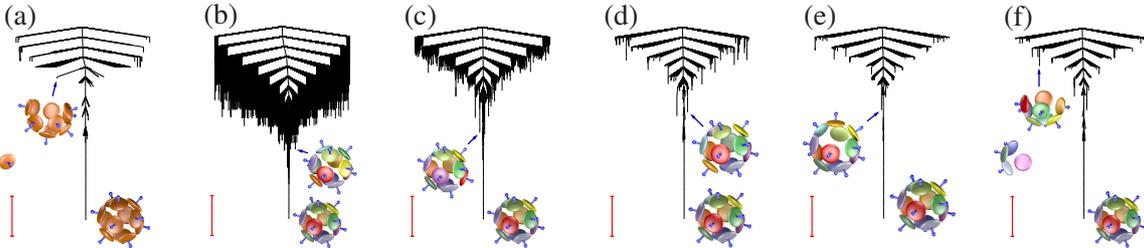


Figure 2. Disconnectivity graphs for addressable $PY_{12}^{I_h}$ shells. $\epsilon_{NN}^{\text{att}} = \epsilon_{NN}^{\text{rep}} = \epsilon_{NNN}^{\text{rep}} = 1$ in each case, except for (f) where $\epsilon_{NN}^{\text{rep}} = 1.5$. (a) $\epsilon_{NNN}^{\text{att}} = 1.00$, (b) $\epsilon_{NNN}^{\text{att}} = 0.75$, (c) $\epsilon_{NNN}^{\text{att}} = 0.50$, (d) $\epsilon_{NNN}^{\text{att}} = 0.25$, (e) $\epsilon_{NNN}^{\text{att}} = 0.00$, (f) $\epsilon_{NNN}^{\text{att}} = 0.00$. The scale bar corresponds to one unit of potential energy and is the same height for each panel. All permutation-inversion isomers are lumped together in panel (a), which corresponds to the usual PY potential. For different formulations of the addressable potential in the other panels enantiomers are paired in each branch.

The disconnectivity graphs for $LJ_{13}^{I_h}$ and $PY_{12}^{I_h}$ indicate that the efficiency of relaxation to the target addressable cluster will increase systematically as $\epsilon_{NNN}^{\text{att}}$ decreases from one to zero (Figures 1 and 2), as the next-nearest neighbour attractions are progressively turned off. The landscapes have maximal bias towards the target when the next-nearest-neighbour attraction is turned off completely, with $\epsilon_{NNN}^{\text{att}} = 0$ (Figures 1e and 2e). The simpler appearance for $\epsilon_{NNN}^{\text{att}} = 1$ in Figures 1a and 2a occurs because all the permutation-inversion isomers of the global minimum reference structure are exactly degenerate, and these structures are collected together in a single branch of the disconnectivity graph. For $\epsilon_{NNN}^{\text{att}} = 0.75$

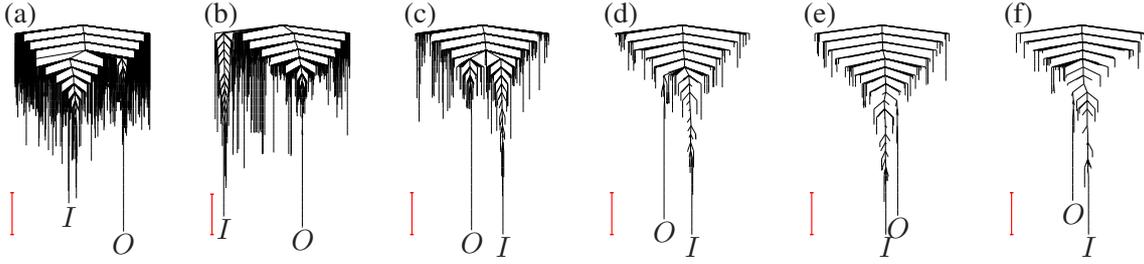


Figure 3. Disconnectivity graphs for doubly addressable LJ_{38} clusters. The two reference structures are the global minimum, with point group O_h , and the second-lowest minimum based on an incomplete Mackay icosahedron [16] with point group C_{5v} . $\epsilon_{\text{NN}}^{\text{att}} = \epsilon_{\text{NN}}^{\text{rep}} = \epsilon_{\text{NNN}}^{\text{rep}} = 1$ in each case, except for (f) where $\epsilon_{\text{NN}}^{\text{rep}} = 1.5$. (a) $\epsilon_{\text{NNN}}^{\text{att}} = 1.00$, (b) $\epsilon_{\text{NNN}}^{\text{att}} = 0.75$, (c) $\epsilon_{\text{NNN}}^{\text{att}} = 0.50$, (d) $\epsilon_{\text{NNN}}^{\text{att}} = 0.25$, (e) $\epsilon_{\text{NNN}}^{\text{att}} = 0.00$, (f) $\epsilon_{\text{NNN}}^{\text{att}} = 0.00$. The scale bar corresponds to one unit of potential energy and is the same height for each panel. All permutation-inversion isomers are lumped together in panel (a), which corresponds to the usual Lennard-Jones potential. For different formulations of the addressable potential in the other panels enantiomers are paired in each branch.

in Figures 1b and 2b this degeneracy is broken, but the energy separation between correctly and wrongly addressed icosahedra is relatively small compared to the results for $\epsilon_{\text{NNN}}^{\text{att}} = 0$ in the (e) panels. The same trends appear for the doubly-addressable potential defined for LJ_{38} with respect to competing close-packed (O) and icosahedral (I) structures (Figure 3), where specific permutational isomers of competing close-packed and icosahedral morphologies are both stabilised with respect to the rest of configuration space. The results for aggregates of multiple copies of the target monomers discussed below all correspond to $\epsilon_{\text{NNN}}^{\text{att}} = 0$.

The present formulation is consistent with current approaches for designing and modelling DNA-coated addressable building blocks [57], in which undesired pair contacts are purely repulsive. We also keep the nearest-neighbour interaction strengths the same for all particles, in line with previous studies [9], leaving the relative value of the non-nearest-neighbour attraction and the repulsion between particles with the same address as the principal adjustable parameters.

The number of local minima decreases with $\epsilon_{\text{NNN}}^{\text{att}}$, as higher energy minima merge with transition states via fold catastrophes [20]. An example for a transformation between the target $LJ_{13}^{I_h}$ icosahedron and an icosahedral structure with misaddressed atoms is shown in Figure 4. In the limit where the higher energy minimum disappears, the fold ratio, which relates the curvature, the path length, and the barrier height, tends smoothly to unity (Figure 4b) as predicted [20].

3.2. Hierarchical assembly of multiple copies of target monomers

Having identified parameters for which the specific target permutational isomer constitutes a particularly low-lying global minimum, we investigated the structure of aggregates for the two types of icosahedral target monomers. Since the global minima for these aggregates will be composed of correctly addressed target monomers, we can explore the relevant configuration space efficiently by moving the monomers as local rigid bodies [58]. Choosing $\epsilon_{\text{self}}^{\text{att}} = 0$ with $\epsilon_{\text{self}}^{\text{rep}} = 10^4$ for the $LJ_{13}^{I_h}$ clusters and $\epsilon_{\text{self}}^{\text{rep}} = 1$ for the $PY_{12}^{I_h}$ shells produces aggregates composed of interacting but distinguishable target monomers. The structure (relative positions of each addressable particle) in each target monomer corresponds to

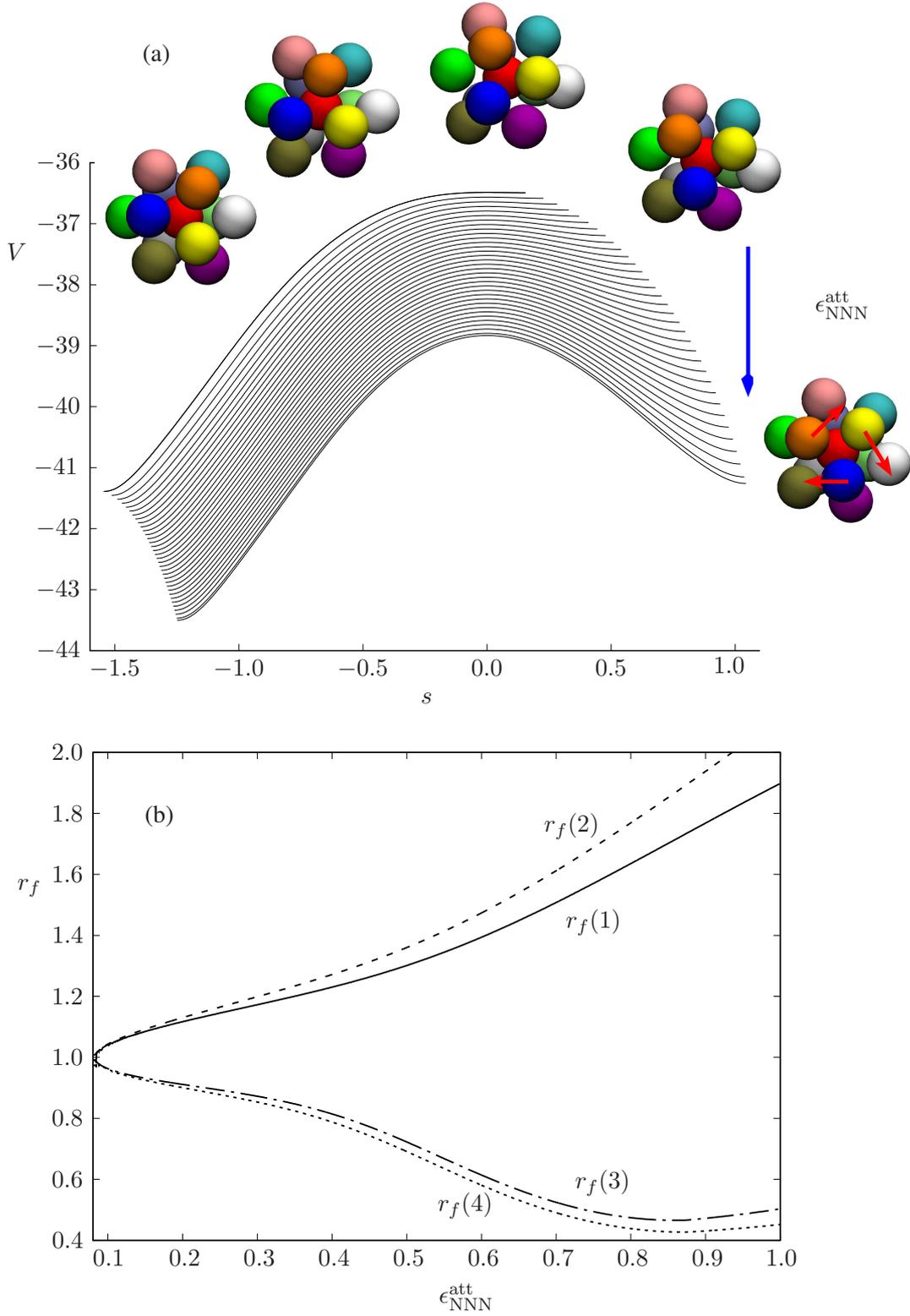


Figure 4. (a) Potential energy, V , as a function of integrated path length, s , for one pathway of the addressable $\text{LJ}_{13}^{I_h}$ cluster as a function of $\epsilon_{\text{NNN}}^{\text{att}}$, corresponding to the fold ratio illustrated in (b). The mechanism is a localised rotation of three atoms in a triangular face, which are misaddressed in the higher energy minimum. Six NNN contacts change to NN for the reference structure in this path, and the energy difference between the two minima scales accordingly, until the misaddressed structure disappears via a fold catastrophe as $\epsilon_{\text{NNN}}^{\text{att}}$ decreases. The envelope of $V(s)$ curves corresponds to regular spacings of $\epsilon_{\text{NNN}}^{\text{att}}$ between 0.70 and 0.08. (b) Fold ratio, r_f , as a function of $\epsilon_{\text{NNN}}^{\text{att}}$ for the pathway of the addressable $\text{LJ}_{13}^{I_h}$ cluster illustrated in (a). Four alternative formulations of r_f are compared, all of which converge smoothly to unity, as predicted by catastrophe theory [20]. $r_f(1)$ and $r_f(2)$ employ the modulus of the negative curvature for the unique normal coordinate with imaginary frequency in the transition state. $r_f(3)$ and $r_f(3)$ are calculated with the curvature of the lowest non-zero frequency normal mode in the higher energy minimum. $r_f(1)$ is slightly less than $r_f(2)$ because it is calculated from the integrated path length rather than the minimum distance between configurations, and similarly for $r_f(3)$ and $r_f(4)$. In the limit of small path length the formulations are all equivalent.

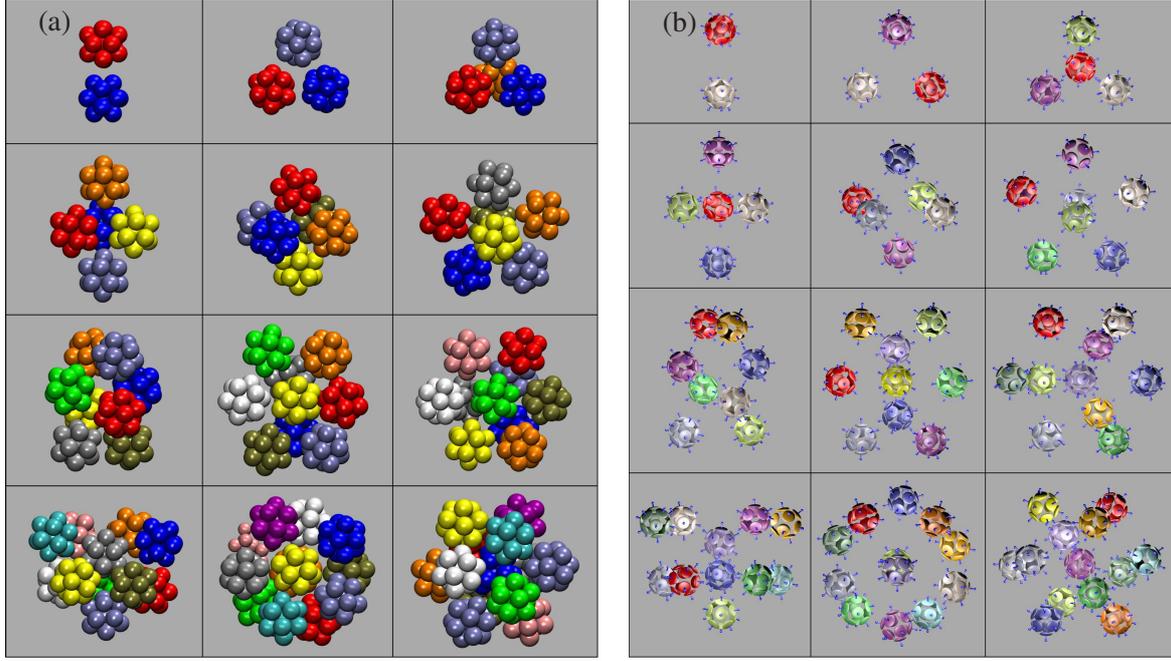


Figure 5. (a) Lowest minima located for 2 to 13 copies of the addressable $\text{LJ}_{13}^{I_h}$ cluster referenced to the 13-atom Mackay icosahedron. $\epsilon_{\text{NN}}^{\text{rep}} = \epsilon_{\text{NN}}^{\text{att}} = \epsilon_{\text{NNN}}^{\text{rep}} = 1$, $\epsilon_{\text{NNN}}^{\text{att}} = 0$, $\epsilon_{\text{self}}^{\text{rep}} = 10^4$, $\epsilon_{\text{self}}^{\text{att}} = 0$. The repulsion between particles with the same address has been chosen sufficiently large for the distinct 13-atom icosahedra to be easily distinguished, and the target monomers are coloured accordingly. (b) Lowest minima located for 2 to 13 copies of the addressable $\text{PY}_{12}^{I_h}$ cluster referenced to the $T = 1$ capsid shell with $\epsilon_{\text{NN}}^{\text{rep}} = \epsilon_{\text{NN}}^{\text{att}} = \epsilon_{\text{NNN}}^{\text{rep}} = 1$, $\epsilon_{\text{NNN}}^{\text{att}} = 0$, $\epsilon_{\text{self}}^{\text{rep}} = 1$, $\epsilon_{\text{self}}^{\text{att}} = 0$. Each icosahedral shell is coloured separately; alternative representations with particles coloured according to their address are shown in Supplementary Figure S1.

the icosahedral global minima presented in Figures 1e and 2e, with small perturbations caused by the interactions between the monomers.

The global minima for $\left[\text{LJ}_{13}^{I_h}\right]_M$ with $2 \leq M \leq 13$ are shown in Figure 5a. The positions defined by the target monomers define well-known structures ranging through the tetrahedron, trigonal bipyramid and octahedron for $M = 4, 5, 6$, up to an icosahedral arrangement for $M = 13$.

Figure 5b shows global minima located for $\left[\text{PY}_{12}^{I_h}\right]_M$, with $2 \leq M \leq 13$. Starting the global optimization from a random configuration, the discoids ($12 \times M$ in total) assemble into correctly addressed 12-particle icosahedral shells, which in turn form the same overall morphologies as the 13-particle addressable LJ icosahedra, but with larger separation between the component monomers, due to the repulsive apex sites pointing outwards from each addressable particle. Alternative representations with particles coloured according to their address are shown in Supplementary Figure S1.

Two alternative representations of $\left[\text{LJ}_{13}^{I_h}\right]_{13}$ are shown in Supplementary Figure S2, one with the target monomers coloured separately, the other with the addressable atoms colour coded and the central atom magnified to highlight the structure. The addressable colouring scheme is adopted for larger clusters in Figure 6. Here we illustrate the lowest minima obtained for aggregates of 55, 147, 309, and 561 target monomers, which correspond to the sizes where complete Mackay icosahedra [16] exist. In each case global optimisation runs were initiated from structures with the target monomers placed on a suitably scaled icosahedral template. Steps were taken using a local rigid body scheme [58] for the target

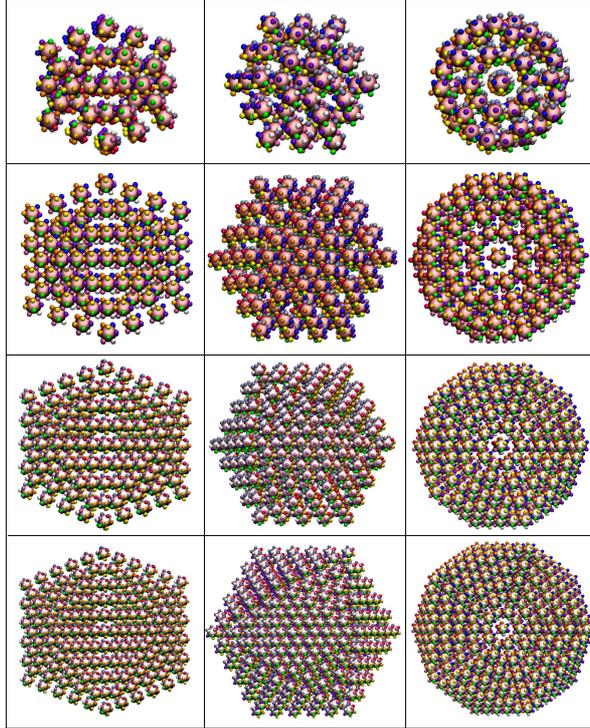


Figure 6. Lowest minima located for $[LJ_{13}^{Ih}]_{55}$, $[LJ_{13}^{Ih}]_{147}$, $[LJ_{13}^{Ih}]_{309}$, $[LJ_{13}^{Ih}]_{561}$ (top to bottom) with $\epsilon_{NN}^{\text{rep}} = \epsilon_{NN}^{\text{att}} = \epsilon_{NNN}^{\text{rep}} = 1$, $\epsilon_{NNN}^{\text{att}} = 0$, $\epsilon_{\text{self}}^{\text{rep}} = 10^4$, $\epsilon_{\text{self}}^{\text{att}} = 0$. The 13 different particles with distinct addresses are coloured consistently, with the central particle magnified to highlight the overall structure. The three views correspond to approximate 2-fold, 3-fold and 5-fold axes (left to right).

monomers, with full relaxation of all atomic degrees of freedom for the lowest minima. At each size the predicted global minimum is based on underlying icosahedral packing. For $[LJ_{13}^{Ih}]_{55}$ and $[LJ_{13}^{Ih}]_{147}$ additional basin-hopping runs were performed from randomly positioned monomers, to check whether the same putative global minimum was located. Eleven out of seventeen runs of 20,000 steps for $[LJ_{13}^{Ih}]_{55}$, and four out of seventeen runs of 15,000 steps for $[LJ_{13}^{Ih}]_{147}$ located the same lowest minimum. The remaining runs succeeded in finding a low-lying minimum with one of the vertex LJ_{13}^{Ih} clusters displaced to a surface site. None of the runs initiated from randomised starting geometries produced a lower energy, providing strong evidence that the structures based on hierarchical icosahedral ordering are indeed the global minima. Only minor local reorientations were observed when the lowest minima identified with rigid target monomers were fully relaxed. The energetic ordering was unchanged on relaxation for all the aggregates.

The hierarchical aggregates of addressable PY particles behave similarly to the addressable LJ clusters. Supplementary Figure S3 shows the lowest energy structure found for $[PY_{12}^{Ih}]_{55}$ by basin-hopping starting with the shells at the vertices of a Mackay icosahedron but randomly oriented.

Experimental realisation of the design principles that emerge from the present analysis will require synthesis of particles capable of matching the addressable interactions. In particular, the potential between particles with the same address will need to be unfavourable, and the attraction between wrongly addressed particles will need to be systematically weaker. The self-avoidance term $\epsilon_{\text{self}}^{\text{rep}}$ is important

in this formulation for hierarchical assembly if the addressable particles interact isotropically. Increasing the magnitude of this term decreases the effective attraction between pairs of icosahedra, therefore widening the energetic gap between the two levels of hierarchical assembly. Encoding such a selective repulsive interaction between isotropic building blocks of the same type represents a significant experimental challenge. We speculate that a design incorporating a charge on the central particle of the $LJ_{13}^{I_h}$ cluster would make it possible to tune the optimal interaction distance between interacting icosahedra. Employing dipolar particles for the outer shell might reinforce this effect. The overall repulsive interactions would further separate the energetic hierarchy involved in the self-assembly process. We plan to explore such alternative formulations in future work.

If the addressable particles have Janus character, as for the $PY_{12}^{I_h}$ clusters, the hierarchical character of the landscape for aggregates arises naturally from the difference between intra-shell and inter-shell forces.

4. Conclusion

For both potentials considered here, aggregation of the monomers composed of addressable particles is hierarchical, since the overall interactions *within* the target monomer are significantly stronger than those *between* monomers. The addressable formulation provides a simple and convenient way to encode hierarchical self-assembling behaviour. We have also shown how the complexity and frustration of the energy landscape for target monomer clusters decreases as attraction between incorrectly addressed particles is reduced. We found that adding a single self-repulsion term between particles in different monomers with the same address contributes to the hierarchical organization of the superclusters corresponding to aggregates of the monomer, and it is possible to vary the magnitude of this parameter to tune hierarchical behaviour. By identifying minimal conditions on the interparticle forces, we hope our results can guide future experiments to realise novel structures with technologically important functionality.

Acknowledgements

SNF acknowledges the grant of the Romanian National Authority for Scientific Research and Innovation, CNCS–UEFISCDI, project number PN-II-RU-TE-2014-4-1176. DJW and RGM gratefully acknowledge support from the EPSRC.

Conflict of interest

There are no conflicts to declare.

References

- [1] S.Y. Park, A.K. Lytton-Jean, B. Lee, S. Weigand, G.C. Schatz and C.A. Mirkin, *Nature* **451** (7178), 553 (2008).
- [2] W.B. Rogers and V.N. Manoharan, *Science* **347** (6222), 639–642 (2015).
- [3] Y. Ke, L.L. Ong, W.M. Shih and P. Yin, *Science* **338**, 1177–1183 (2012).
- [4] A. Reinhardt and D. Frenkel, *Phys. Rev. Lett.* **112**, 238103 (2014).
- [5] W.M. Jacobs, A. Reinhardt and D. Frenkel, *Proc. Natl. Acad. Sci. USA* **112**, 6313–6318 (2015).
- [6] G. Meng, N. Arkus, M.P. Brenner and V.N. Manoharan, *Science* **327**, 560–563 (2010).
- [7] J.C. Crocker, *Science* **327**, 535–536 (2010).
- [8] D.J. Wales, *ChemPhysChem* **11**, 2491–2494 (2010).
- [9] S. Hormoz and M.P. Brenner, *Proc. Nat. Acad. Sci. USA* **108**, 5193 (2011).
- [10] D. Frenkel and D.J. Wales, *Nature Materials* **10**, 410–411 (2011).
- [11] F. Calvo, J.P.K. Doye and D.J. Wales, *Nanoscale* **4**, 1085 – 1100 (2012).
- [12] D.J. Wales, *J. Chem. Phys.* **146** (5), 054306 (2017).
- [13] W.M. Jacobs and D. Frenkel, *J. Am. Chem. Soc.* **138** (8), 2457–2467 (2016).
- [14] A. Murugan, J. Zou and M.P. Brenner, *Nat. Commun.* **6**, 6203 (2015).
- [15] A.V. Tkachenko, *Phys. Rev. Lett.* **106** (25), 255501 (2011).
- [16] A.L. Mackay, *Acta Cryst.* **15**, 916 (1962).
- [17] D.L.D. Caspar and A. Klug, *Cold Spring Harbour, Symp. Quant. Biol.* **27**, 1 (1962).
- [18] O.M. Becker and M. Karplus, *J. Chem. Phys.* **106**, 1495–1517 (1997).
- [19] D.J. Wales, M.A. Miller and T.R. Walsh, *Nature* **394**, 758–760 (1998).
- [20] D.J. Wales, *Science* **293**, 2067–2069 (2001).
- [21] J.E. Jones and A.E. Ingham, *Proc. R. Soc. A* **107**, 636–653 (1925).
- [22] Y. Ueda, H. Taketomi and N. Gō, *Biopolymers* **17**, 1531 (1978).
- [23] R.B. Best, Y.G. Chen and G. Hummer, *Structure* **13**, 1755–1763 (2005).
- [24] K.i. Okazaki, N. Koga, S. Takada, J.N. Onuchic and P.G. Wolynes, *Proc. Nat. Acad. Sci. USA* **103**, 11844–11849 (2006).
- [25] S. Franz and G. Parisi, *Phys. Rev. Lett.* **79**, 2486–2489 (1997).
- [26] L. Berthier and R.L. Jack, *Phys. Rev. Lett.* **114**, 205701 (2015).
- [27] R.L. Jack and J.P. Garrahan, *Phys. Rev. Lett.* **116**, 055702 (2016).
- [28] C.J. Fullerton and R.L. Jack, *Phys. Rev. Lett.* **112**, 255701 (2014).
- [29] R.L. Jack and C.J. Fullerton, *Phys. Rev. E* **88**, 042304 (2013).
- [30] L. Berthier and W. Kob, *Phys. Rev. E* **85**, 011102 (2012).
- [31] W. Kob and L. Berthier, *Phys. Rev. Lett.* **110**, 245702 (2013).
- [32] S. Karmakar and G. Parisi, *Proc. Natl Acad. Sci. USA* **110**, 2752–2757 (2013).
- [33] C. Cammarota and G. Biroli, *ArXiv e-prints* (2011).
- [34] C. Cammarota and G. Biroli, *Proc. Natl Acad. Sci. USA* **109**, 8850–8855 (2012).
- [35] C. Cammarota and G. Biroli, *J. Chem. Phys.* **138**, 12A547 (2013).

- [36] K. Kim, *Europhysics Lett.* **61** (6), 790–795 (2003).
- [37] M. Ozawa, W. Kob, A. Ikeda and K. Miyazaki, *PNAS* **112** (22), 6914–6919 (2015).
- [38] S. Torquato, *Soft Matter* **5** (6), 1157–1173 (2009).
- [39] M.C. Rechtsman, F.H. Stillinger and S. Torquato, *Phys. Rev. E* **75**, 031403 (2007).
- [40] E. Marcotte, F.H. Stillinger and S. Torquato, *J. Chem. Phys.* **138** (6), 061101 (2013).
- [41] E. Marcotte, F.H. Stillinger and S. Torquato, *Soft Matter* **7**, 2332–2335 (2011).
- [42] M.C. Rechtsman, F.H. Stillinger and S. Torquato, *Phys. Rev. Lett.* **95**, 228301 (2005).
- [43] E. Marcotte, F.H. Stillinger and S. Torquato, *J. Chem. Phys.* **134** (16), 164105 (2011).
- [44] L. Paramonov and S.N. Yaliraki, *J. Chem. Phys.* **123** (19), 194111 (2005).
- [45] S.N. Fejer, D. Chakrabarti and D.J. Wales, *ACS Nano* **4**, 219–228 (2010).
- [46] S.N. Fejer, D. Chakrabarti and D.J. Wales, *Soft Matter* **7**, 3553–3564 (2011).
- [47] C.J. Forman, S.N. Fejer, D. Chakrabarti, P.D. Barker and D.J. Wales, *J. Phys. Chem. B* **117** (26), 7918–7928 (2013).
- [48] S.W. Olesen, S.N. Fejer, D. Chakrabarti and D.J. Wales, *RSC Adv.* **3**, 12905–12908 (2013).
- [49] D.J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- [50] D.J. Wales and H.A. Scheraga, *Science* **285**, 1368–1372 (1999).
- [51] D.J. Wales and J.P.K. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- [52] S.A. Trygubenko and D.J. Wales, *J. Chem. Phys.* **120**, 2082–2094 (2004).
- [53] G. Henkelman, B.P. Uberuaga and H. Jónsson, *J. Chem. Phys.* **113**, 9901–9904 (2000).
- [54] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **113**, 9978–9985 (2000).
- [55] L.J. Munro and D.J. Wales, *Phys. Rev. B* **59**, 3969–3980 (1999).
- [56] Y. Zeng, P. Xiao and G. Henkelman, *J. Chem. Phys.* **140** (4), 044115 (2014).
- [57] Z. Zeravcic, V.N. Manoharan and M.P. Brenner, *Proc. Natl. Acad. Sci. USA* **111** (45), 15918–15923 (2014).
- [58] H. Kusumaatmaja, C.S. Whittleston and D.J. Wales, *J. Chem. Theor. Comput.* **8** (12), 5159–5165 (2012).