The equilibrium structure of self-assembled protein nano-cages

Journal:	*Nanoscale*
Manuscript ID	NR-ART-09-2018-007202.R1
Article Type:	Paper
Date Submitted by the Author:	25-Oct-2018
Complete List of Authors:	Panahandeh, Sanaz; University of California Riverside, Physics Li, Siyu; University of California Riverside, Physics Zandi, Roya; University of California Riverside, Physics

ROYAL SOCIETY
OF CHEMISTRY

Journal Name

ARTICLE TYPE

The equilibrium structure of self-assembled protein nano-cages[†]

Sanaz Panahandeh,[a] Siyu Li[a] and Roya Zandi[a]

Understanding how highly symmetric, robust, monodisperse protein nano-cages self-assemble can have major applications in various areas of bio-nanotechnology, such as drug delivery, biomedical imaging and gene therapy. We develop a model to investigate the assembly of protein subunits into the structures with different size and symmetry. Using Monte Carlo simulation, we obtain the global minimum energy structures. Our results suggest that the physical properties including the spontaneous curvature, flexibility and bending rigidity of coat proteins are sufficient to predict the size, symmetry and shape selectivity of the assembly products. Further, on a thermodynamic basis, we discuss the polymorphism of nano-cages observed in assembly experiments.

Introduction

Self-assembly of monodispersed protein cages is ubiquitous in nature. Because of their biocompatibility, stability and low toxicity, protein cages have important roles in many biological processes, medicine and bio-nanotechnology. Examples of protein cages include platonic hydrocarbons, heat shock proteins, ferritins, carboxysomes, silicages, multicomponent ligand assemblies, clathrin vesicles and virus shells, to name a few[1–5]. The protein shells are necessary for both protection and delivery of various cargos in biological systems. For instance, ferritin stores iron and exists in almost every living organism.

Among all biological entities, viruses in particular have optimized the feat of packaging of genetic materials and other anionic cargos into a protein shell called the capsid, recognized as one of the most efficient nano-containers for trafficking genetic material in nature[6,7]. Most protein cages self-assemble from a large number of one or a few different types of protein subunits into complex supramolecular structures with diameters ranging from 10 to 500 nm[8]. Quite remarkably under many circumstances, viruses spontaneously assemble *in vitro* from protein building blocks into highly symmetric shells[9,10]. Most spherical viruses adopt structures with icosahedral symmetry[11–13] characterized by a structural index T number, which assume only certain integers (1, 3, 4, 7, ...)[14]. The number of protein subunits in icosahedral shells is often 60 times the T-number.

Other protein cages can adopt several other symmetric structures. For example, clathrin shells form icosahedral structures in
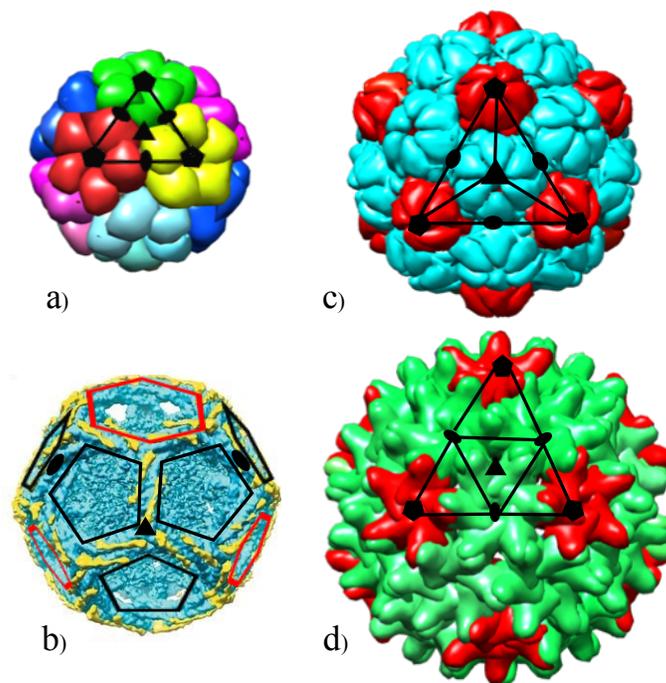
Fig. 1 Structures of some protein cages. (a) Lumazine synthase is an enzyme with icosahedral symmetry ($T = 1$) constructed of 60 identical protein subunits. The colors are added to highlight each pentamer. (b) Mini-coat has tetrahedral symmetry[15]. The two-fold and three-fold symmetry axes are marked with small black ovals and triangles respectively. (c) Encapsulin from M. xanthus with $T = 3$ structure is made of 180 identical protein subunits. The position of two, three and five-fold symmetry axes are marked in the picture. (d) Hepatitis B virus, a $T = 4$ structure. The darker color (red) in (c) and (d) are pentamers. All the structures except (b) are reproduced using UCSF Chimera packages (http://www.rbvi.ucsf.edu/chimera).

[a] Department of Physics and Astronomy, University of California, Riverside, California 92521, USA; E-mail: spana002@ucr.edu

addition to many other symmetric shells[16,17] depending on the size of their cargo. Nevertheless protein cages with icosahedral symmetry are by far the most abundant in nature. Figure 1 illustrates the structure of Lumazine synthase with $T = 1$ symmetry[18], a Clathrin shell with tetrahedral symmetry[15], an encapsulin nanocompartment from M. xanthus with $T = 3$ structure[19] and the Hepatitis B virus (HBV) capsid with $T = 4$ symmetry.

Despite the abundance of protein shells in nature, the role of building blocks and the factors contributing to the stability, size and shape selectivity of nanostructures are not well-understood. To this end, there is a precedent need to take a bottom-up approach and to understand at the fundamental scale the impact of building blocks on the design and formation of functional nanoshells.

Extensive work has explored the effect of spontaneous radius of curvature (dihedral angle) of building blocks on the equilibrium structure of protein cages[20]. For instance, using the Monte Carlo (MC) simulations, Chen *et al.* studied the self-assembly of attractive cone-shaped particles into different structures[21]. They obtained a sequence of clusters and found that the symmetry and stability of formed structures depend on the cone angle or the preferred angle between subunits. Similar sequence of structures was obtained with attractive spherical particles but under certain convexity constraints, equivalent of changing the preferred dihedral angle between subunits[22].

The simple case of N spherical colloids or circular disks interacting through Lennard-Jones potential constrained to move on the surface of a sphere also shows that the equilibrium structure of shells depends on the number of building blocks and the preferred angle between disks or Lennard-Jones particles[23]. As the preferred angle between disks or colloids changes, structures with different size and symmetries form.

More recently, Paquay *et al.* studied the equilibrium structures of interacting Morse particles residing on the surface of a sphere and found similar structures and magic numbers as observed in the case of LJ particles[24]. Nevertheless, the impact on the equilibrium structures of the mechanical properties of building blocks including flexibility and bending rigidity have not previously been studied. While the dynamical structures of protein shells under non-equilibrium conditions as a function of bending rigidity and stretching modulus of building blocks have been thoroughly investigated in Ref.[25], due to irreversible steps in the shell growth, the structures obtained in those simulations might be completely far from equilibrium.

In this paper we investigate the equilibrium structure of nanoshells and the important factors contributing to their stability and symmetry. Using MC simulations combined with the bond flipping method[26,27] we study the structure of protein cages as a function of the spontaneous curvature as well as stretching and bending rigidity of building blocks, advancing our knowledge for producing high yield nano-cages with specific size and shape.

While spontaneous curvature is an important factor in defining the size of the shell, we find that the flexibility and bending rigidity of building blocks can completely modify the size and final symmetry of the shells. Quite interestingly the sequence of clusters or magic numbers and their associated shells obtained in our equilibrium studies, coincide not only with the structure of viruses displaying icosahedral symmetry but with other non-icosahedral protein cages observed in other systems such as clathrin shells.

Furthermore, we find that there are striking similarities between the minimum energy structure phase diagram and the one obtained through irreversible growth[25], as a function of the mechanical properties and spontaneous curvature of building blocks. Both diagrams display the same symmetry for almost the same regions of the parameter space. We emphasize that the location of pentamers in a shell defines its symmetry. If the position of a pentamer in a symmetric shell is slightly displaced, the symmetry will be broken. The fact that despite the irreversible steps, pentamers form in the "right" positions, preserving the symmetry during the irreversible shell growth, is quite unexpected. The similarities between irreversible and equilibrium structures in this paper can be explained at least in part by the recent work of Li *et al.*. Using continuum elasticity theory, they have shown that as an elastic shell grows, there are regions on the spherical cap that strongly adsorb disclinations (pentamers), *i.e.*, there is a high affinity for the formation of disclinations at certain locations as the cap grows. We will discuss this effect more in the conclusion of the paper.

We find some differences between two phase diagrams too. A few symmetric structures grown in irreversible simulations[25] do not constitute the minimum free energy structures. Furthermore, we obtain additional symmetric structures in the equilibrium simulations, which were not observed in the growth simulations under irreversible conditions.

We note here that in our MC simulations the assembled shells are obtained under the condition that the thermal fluctuations are negligible compared to the other energies in the system, and as such the equilibrium structures coincide with the minimum energy structures.

It is also worth mentioning that it is now widely accepted that the preferred curvature and mechanical properties of subunits depend on the solution conditions such as pH and salt concentration[28]. The interplay of protein geometry, repulsive electrostatic and attractive hydrophobic interactions define the equilibrium properties (bending and stretching moduli and spontaneous curvature) of subunits, nevertheless no systematic experimental data are known for these parameters. In addition to solution conditions, mutations can also affect the physical properties of protein subunits[29], enabling us to test several theories in this paper. Understanding the role of stiffness and preferred curvature of building blocks could lead to generation of a range of new materials and novel structures.

Method

To study the equilibrium structures, we consider stretchable equilateral triangular subunits, representing building blocks of protein cages, as illustrated in Fig. 1a and d. Triangular subunits are suitable to describe the structure of protein cages as they form hexagonal sheets in flat space and are able to build a spherical mesh with at least 12 five coordinated lattice points (pentagons). The total energy of a triangular shell is the sum of the stretching and bending energies[30,31]. The stretching energy of triangular network can simply be defined by a harmonic potential summed

over all triangles,

$$E_s = \sum_i \sum_{a=1}^{3} \frac{k_s}{2} (b_i^a - b_0)^2 \tag{1}$$

with i the triangular subunit index, b_0 the equilibrium length of the edges, and b_i^a the length of the ath edge in the ith subunit. While the stretching energy is related to the deformation of subunits from their equilateral shapes, the bending energy corresponds to the deviation of the dihedral angle between adjacent subunits from the preferred one. The bending energy is obtained by summing over all pairs of triangular subunits that share an edge and can be written as,

$$E_b = \sum_{<ij>} k_b(1 - \cos(\theta_{ij} - \theta_0)) \tag{2}$$

with $<ij>$ the index pairs of neighboring subunits, k_b the torsional spring constant, and θ_0 the preferred dihedral angle between two subunits. The preferred dihedral angle and spontaneous radius of curvature are related through $\sin(\theta_0/2) = (12R_0^2/b_0^2 - 3)^{-1/2}$ with R_0 the spontaneous radius of curvature. The angle θ_{ij}, where $\cos\theta_{ij} = \hat{n}_i \cdot \hat{n}_j$, is between the unit normal vectors $\hat{n}_i$ and $\hat{n}_j$ of the two adjacent subunits i and j sharing an edge.

Equations 1 and 2 reveal the presence of two important dimensionless parameters, the spontaneous radius of curvature R_0/b_0 and the Foppl von Karman (FvK) number

$$\gamma = k_s b_0^2 / k_b, \tag{3}$$

which indicates the relative difficulty of deforming an equilateral triangular subunit compared to changing the dihedral angle between two adjacent subunits away from the preferred one. We note that both dimensionless parameters are normalized with respect to the size of the subunits b_0.

To obtain the lowest-energy configurations we employ a series of simulated annealing MC simulations[32]. We start from a triangulated spherical mesh with a random distribution of N_v vertices. Each MC step consists of N_v attempted bond movings, which involves removing and reattaching the edge connecting two vertices of two neighboring triangles such that the two vertices which were not connected before, they will be linked by an edge after the flip, as is shown in Fig. 2. Detachment and reconnection of the bonds are such that the total number of vertices N_v, subunits n_s and edges in the shell remain constant. Since our goal is to obtain the global minimum energy structure each edge swapping is followed by the shell relaxation during which vertices will move to the positions that minimize the total elastic energy. We employ the BFGS method to relax and minimize the energy of the shell[33]. The probability that the new relaxed structure with the new position of vertices be accepted is $\min(1, e^{(E_{old} - E_{new})/k_B T})$. E_{old} and E_{new} are the energies of the structures before and after the trial edge swapping, respectively. We generate a Markov chain with Boltzmann probabilities by iterating the edge swapping until the energy converges. The edge swapping process is reversible to ensure detailed balance.

We repeat the above simulations with different initial configu-
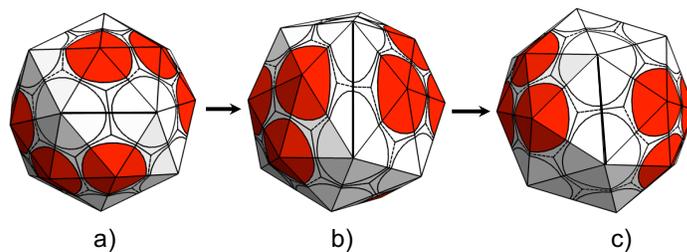
Fig. 2 Bond moving method: (a) The thick black edge between two neighboring triangles is randomly chosen. (b) The black line is removed from its previous position and the two vertices that were not sharing a bond before the swap, are now connected. The darker (red) shades indicate the positions of pentamers and the white ones correspond to hexamers. By moving the bond from (a) to (b), the position of pentamers and hexamers are changed. (c) The system is energetically relaxed now after the swap.

rations many times. To avoid local minimum free energy traps, we employ simulated annealing with both linear and non-linear schedules[21]. The results show no dependence on the cooling path as long as the cooling rate is very slow. We note again that since the thermal fluctuations are negligible compared to the other energies in the system, the structures we obtain correspond to the global minimum energy structures.

The above algorithm allows us to successfully change the position of pentamers and hexamers. In other words, during the simulations the location of disclinations is not fixed; they can move, and thus change the structure and symmetry of the shell. We perform MC simulations for all the structures ranging from $N_v = 12$ to 42 corresponding to the shells made of $n_s = 20$ to 80 number of subunits. In nature larger shells need some external help like scaffolding proteins or inner core to form symmetric shells[34], the focus of this paper is, however, on the smaller shells that are able to spontaneously assemble without any core.

Results

We carry out Monte Carlo simulations as described in the previous section for different numbers of subunits n_s. We start with a fixed preferred spontaneous radius of curvature $R_0/b_0 = 1.28$ but different values of γ. The results of the simulations are illustrated in Fig. 3 in the form of a plot of the minimized elastic energy per triangles ε_n (in units of $k_B T$) versus the number of subunits, n_s. The solid light line (green) corresponds to $\gamma = 0.5$, the dashed line to $\gamma = 1$, the dark solid line to $\gamma = 3$ and the dotted line to $\gamma = 8$. We emphasize that since γ is proportional to the ratio of stretching to bending modulus, for larger γs it is difficult to deform the subunits from their equilibrium equilateral shape but rather easy to bend them away from their preferred dihedral angle. For small γs, in contrast, the subunits can be easily deformed but it costs significantly more energy to modify the dihedral angle between the adjacent subunits from the preferred one.

Figure 3 illustrates that there are many local minima but no distinguished global minimum energy for $\gamma = 0.5$. The energy landscape indeed becomes flatter as $\gamma \to 0$. When γ becomes smaller, the subunits becomes more flexible, and thus the size of the shell changes with R_0/b_0. However, as γ becomes larger, the subunits
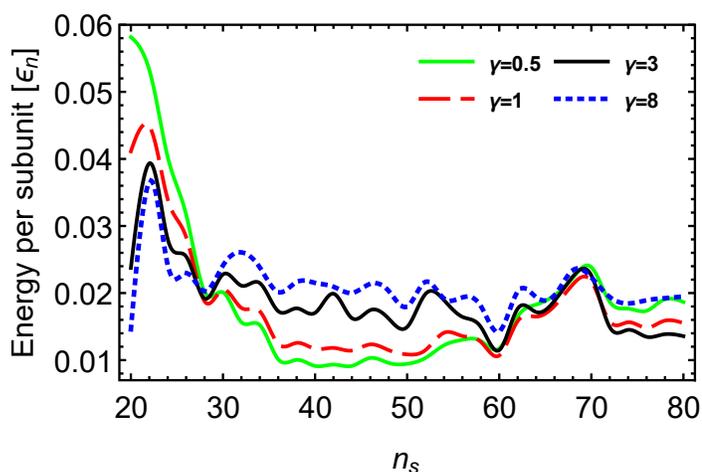
Fig. 3 Energy per subunit versus number of subunits are displayed for $\gamma = 0.5$, 1, 3 and 8. The spontaneous radius of curvature is fixed at $R_0/b_0 = 1.28$. While the minimum energy for $\gamma = 0.5$ is at $n_s = 40$, for $\gamma = 1$, 3 and 8 the equilibrium structure is a $T = 3$ icosahedral structure.

Fig. 4 Plot of energy per subunit versus R_0/b_0 for $\gamma = 0.5$, 1, 3 and 8. Each curve can be divided into different segments. The capital letter at the beginning of each segment (from left to right) indicates the symmetry of the segment. The letter A pertains to the beginning of all curves. The corresponding structures are illustrated in Fig. 5a.

become more rigid and avoid configurations leading to the deformation of equilateral triangles, and as such only the structures with icosahedral symmetry remain the global minimum energy structure. For $\gamma < 1$, other structures with different n_s compete or have lower energies than a $T = 3$ structure.

We next investigate the impact of the spontaneous radius of curvature on the global energy minima of Fig. 3 for various γs. Figure 4 illustrates the plot of energy per subunit versus R_0/b_0 for the global minimum energy structures (Fig. 3) at different γ-values. The curves in Fig. 4 can be divided into different segments, each representing different structure. The capital letter at the beginning of each segment reveals the symmetry and structure of that segment. Note that letter A pertains to the beginning of all curves. For instance, the letter A at the beginning of the thick dotted line shows that for $\gamma = 8$ the global minimum energy structure is a $T = 1$ icosahedral shell when $1 < R_0/b_0 < 1.3$. All the structures corresponding to the capital letters are illustrated in Fig. 5a.

The thick dotted line in Fig. 4 shows that even though the energy per subunit increases as R_0/b_0 increases, $T = 1$ remains the global minimum energy structure till $R_0/b_0 = 1.3$ when the icosahedral $T = 3$ becomes the global minimum energy structure. This effect is also apparent in Fig. S3 †, which is a plot of the number of subunits n_s versus R_0/b_0. There is a big jump in the number of subunits from $n_s = 20$ ($T = 1$) to $n_s = 60$ ($T = 3$) at $R_0/b_0 = 1.3$.

All the above effects can be seen more clearly in Fig. 5b in the form of a "shape" phase diagram of spontaneous radius of curvature R_0/b_0 and γ. Each shaded region in the diagram corresponds to a different shell whose structure and symmetry are illustrated in Fig. 5a. Figure 5b shows that the structures become more sensitive to the spontaneous radius of curvature as γ decreases. For instance, for $\gamma = 0.2$ as the spontaneous radius of curvature varies, we obtain nine different symmetric shells between $R_0/b_0 = 1$ and 1.7, see also Fig. S4 †. However, there are only two different structures at $\gamma = 8$ over a wide range of spontaneous curvature,
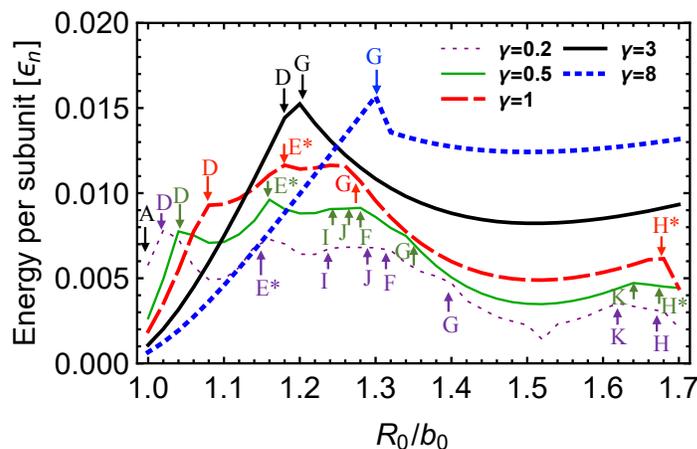
$T = 1$ and $T = 3$. This is basically due to the fact that at larger γs the protein building blocks are stiffer and it is energetically more costly to deform them from their native shape. Since for icosahedral structures most proteins are sitting in equivalent positions, at high γ-values icosahedral structures are the minimum energy structures for the range of the spontaneous curvature studied, as illustrated in Fig. 5b.

The largest shell obtained in Fig. 5b contains $n_s = 80$ triangles corresponding to a $T = 4$ structure for smaller γ-values. Note that at intermediate γs, another equilibrium structure with the same number of subunits as a $T = 4$ shell ($n_s = 80$) but different symmetry exists, which we label it as H^* in Figs. 5a and 5b. While $T = 1$ and $T = 3$ occupy large regions in the equilibrium phase diagram, only a small region belongs to $T = 4$. This is consistent with several previous studies. First of all, a review of literature shows that there are fewer $T = 4$ structures in nature[35,36]. Furthermore, in Refs.[22,24] only the D_{5h} structure mentioned above was observed and no $T = 4$ icosahedral structures appeared in their simulations.

It is now interesting to compare the equilibrium "shape" phase diagram with the diagram obtained through irreversible assembly[25].

Equilibrium versus non-equilibrium
The non-equilibrium or irreversible structures are obtained based on the algorithm used in Ref.[25], and are illustrated in the form of a phase diagram of the dimensionless ratio of bending to stretching modulus γ and the spontaneous radius of curvature R_0/b_0 in Fig. 6. As in the case of the equilibrium phase diagram, each color refers to a different symmetric structure. For completeness, the details of obtaining non-equilibrium or irreversible structures are given in the SI †. Briefly, these structures grow following the local minimum energy pathway. Nevertheless, during the assembly process once a pentamer or hexamer forms, its position is permanently fixed. Due to these irreversible steps, the structures of assembled shells could be completely far from equilibrium. It
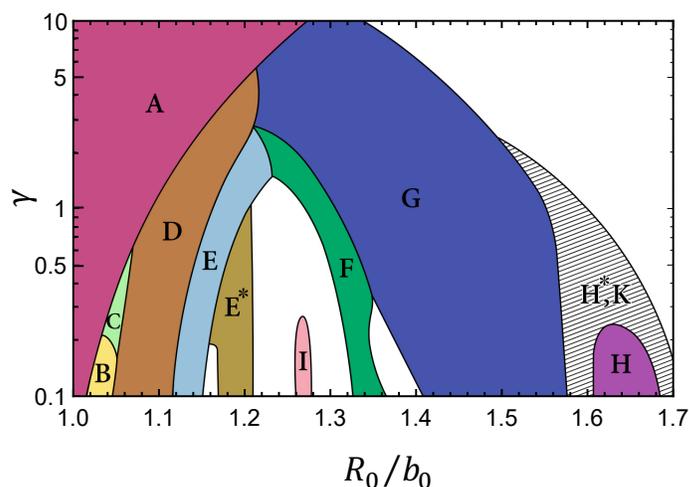
a)

b)
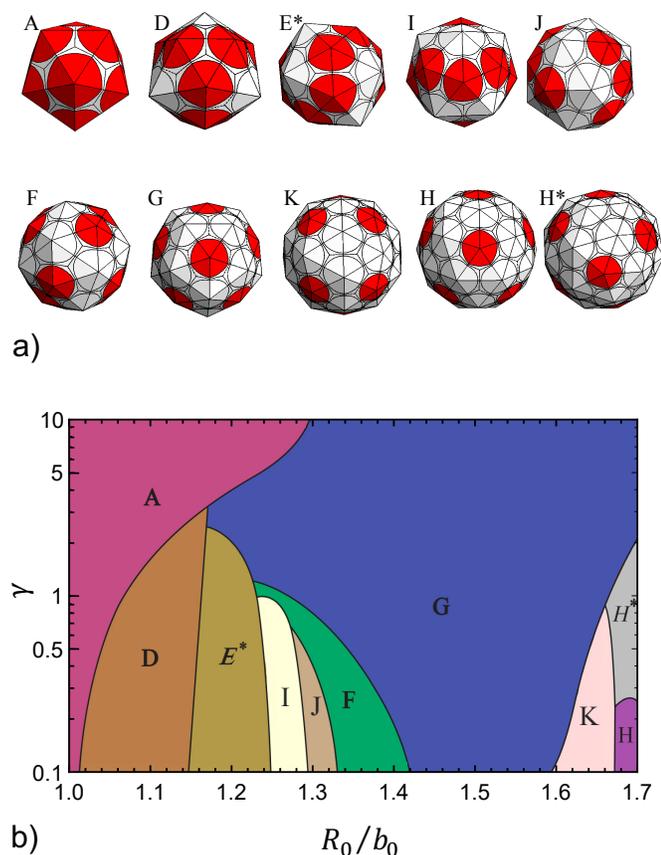
Fig. 5 (a)The equilibrium structures obtained in the simulations corresponding to the labeled regions of the phase diagram illustrated in Figure b. The shells from left to right and top to bottom have $n_s = $ 20, 28, 36, 40, 44, 50, 60, 76, 80 and 80 subunits and symmetries are icosahedral ($T = 1$), tetrahedral, D_2(tennis ball), D_2, D_2, D_3, icosahedral ($T = 3$), tetrahedral, icosahedral ($T = 4$) and D_{5h}, respectively. (b) Phase diagram of the equilibrium structures presenting various shells assembled for different values of γ and R_0/b. Each shaded region corresponds to a single equilibrium shell type . Region A corresponds to a shell with $n_s = 20$. The regions $D - K$ correspond to shells with $n_s = 28, 36, 50, 60, 80, 40, 44$ and 76 subunits. Both H and H^* structures have the same number of subunits $n_s = 80$.

Fig. 6 Phase diagram of the irreversible growth: the dark shaded contiguous regions labeled by letters A through H correspond to regions where only a single type of symmetric shell is assembled. In the hashed region some irregular shells grow in addition to H^* and K structures. The majority of K structures form at the phase boundary between H^* and the adjacent white region. The white areas show the regions in which different types of shells without any specific symmetry are formed.

appears that this "irreversible" algorithm has been successful in explaining many viral structures[30,31,37,38].

We find it quite striking that the shell assembly along the local minimum energy path with the restrictive conditions of irreversible growth leads to the formation of shells almost identical to those obtained in equilibrium studies. These results are quite unexpected considering that the principles of detailed balance is violated in the irreversible growth and as such one would expect a big difference between the two phase diagrams.

Despite the similarities, there are some differences between the two phase diagrams, see Figs. 5b and 6. The shells that only appear in the irreversible phase diagram are illustrated in Fig. 7. Two small regions (structures B and C) in the irreversible phase diagram (Fig. 6) corresponding to $n_s = 24$ and 26 do not constitute the minimum free energy structures. In the equilibrium phase diagram, they are both replaced by the structure D, a clathrin shell, which has $n_s = 28$ and is called mini-coat. The

other clathrin shells, hexagonal barrel (structure E) obtained in the irreversible growth has $n_s = 36$ with D_{6h} dihedral symmetry. The equilibrium structure of the shell with the same $n_s = 36$ has tennis ball symmetry, the structure E^* in Figs. 5b and 6.

The white area in the irreversible phase diagram for $\gamma < 2$ and between $1.21 < R_0/b < 1.3$ ($n_s = 38 - 48$) corresponds to the region where many different types of shells without any specific symmetry are assembled. In contrast, there is no irregular structure in the equilibrium phase diagram, and we find the structures I ($n_s = 40$) and J ($n_s = 44$) with D_2 symmetry in that region.

Moreover, the regions corresponding to G and F structures cover a larger parameter space in the equilibrium phase diagram compared to the irreversible one, revealing the presence of energy barriers as the symmetric shells grow. The energy barrier is between a local minimum energy that the growing shell is trapped into it because of the irreversible assembly path and the global minimum energy.

Furthermore, the irregular shells formed between the G and F structures at lower γ-values in the irreversible phase diagram disappear and are replaced with the F one. The structure K with $n_s = 76$ and tetrahedral symmetry which forms between G and H (or H^*) in the equilibrium phase diagram, appears rarely in the irreversible one. In fact it only assembles at the boundary between the hashed and the white regions (Fig. 6), despite the fact that the K structure is smaller than H or H^*. Last but not least, the structures with icosahedral symmetry cover a wider region in the equilibrium phase diagram.

The largest symmetric shell in Figs. 5b and 6 is $n_s = 80$. Note that in the absence of a cargo which could be a genome, an inner
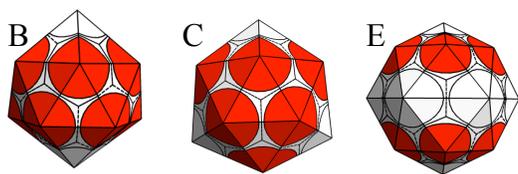
Fig. 7 The symmetric shells growing under irreversible assembly conditions, which do not appear in the equilibrium phase diagram. The shells from left to right have $n_s = 24$, 26 and 36 subunits. The symmetries are D_6, D_3 and D_{6h} respectively. In the irreversible phase diagram (Fig. 6) they are labeled as B, C and E. The E structure has the same number of subunits as the E^* structure but with different symmetry.

shell or scaffolding proteins, at low γ, when we increase the spontaneous radius of curvature only irregular shapes form. However, at large γ-values for $(R_0/b_0 \gg 1)$, we obtain flat sheets or other structures with zero Gaussian curvature[30,31,37,39]. In this paper, we only focus on the assembly of small symmetric shells, as illustrated in Figs. 5b.

Discussion and Conclusion

Despite the wide range of amino acid sequences and folding structures of coat proteins, many protein cages spontaneously self-assemble to form icosahedral or other symmetric structures. This reveals a "universal" behavior among most protein cages in nature. In this paper, using the MC simulation and edge swapping method, we investigated the equilibrium structure of protein cages built from identical subunits. We, in particular, choose triangular subunits as they are a generic choice that applies to the majority of protein nano-cages. Since triangles form a hexagonal lattice in flat space and also assemble to create 12 disclinations for making closed polyhedrons, their behavior is similar to the building blocks of protein cages. While in this paper and many others equilateral triangular subunits have been able to capture the universal behavior of viruses and other proteins cages[30,37–41], we do not expect that the equilateral triangular mesh explains the behavior of systems in which anisotropy of subunits plays an important role in the final structure of protein cages.

Using triangular subunits, we studied the impact of the mechanical properties of building blocks on the symmetry and structure of small protein cages and constructed a phase diagram as a function of the spontaneous curvature of subunits and FvK number (the ratio of stretching to bending modulus), as shown in Fig. 5b. The phase diagram is significantly occupied with icosahedral shells, $T = 1$ and 3, which are common among viruses and many other protein cages. As illustrated in the figure, at low γ, where subunits can deform easily from the equilateral triangle, various structures form as a function of spontaneous curvature. However, by increasing γ, subunits become more rigid and structures with lower symmetries disappear. For instance for low $\gamma = 0.2$, the equilibrium structures are sensitive to the spontaneous radius of curvature and there is a smooth transition from one shell to the next one as illustrated in Fig. S4[†]. Increasing to $\gamma = 8$, the equilibrium structures become less sensitive to R_0/b_0 and only icosahedral structures with $T = 1$ and 3 survive, indicating the robustness of these two structures.

Quite unexpectedly, we found that the equilibrium phase diagram, Fig. 5b, was very similar to the phase diagram obtained under irreversible conditions, see Fig. 6. As explained in the previous section, the irreversible simulations of Ref.[25] were performed following the local minimum energy path but under the condition that once a pentamer or hexamer formed, it could no longer dissociate or move. Since the principles of detailed balance were violated in the irreversible growth simulations, we did not expect to observe such a high degree of similarity between the two phase diagrams, see Figs. 5b and 6.

These results could be explained to some extent with the recent work of Li *et al.* who employed the continuum elasticity theory and studied the assembly pathway of icosahedral shells. They found that as an elastic shell grows, there is a deep potential well attracting pentamers exactly at the locations that will become the vertices of an icosahedron when the shell is complete[38].

Due to the small size and discreteness of the system, the continuum elasticity theory cannot explain the symmetry of the shells observed in Figs. 5a and 7. However, the extensive similarities between equilibrium and irreversible phase diagrams in Figs. 5b and 6 indicate that for the symmetric shells other than icosahedral ones, there are also high-affinity regions for the formation of disclinations at specific locations during the growth process, which leads to the assembly of different types of symmetric shells, depending on the mechanical properties of protein subunits.

It is important to note that we often found one single global minimum energy structure in the phase diagrams presented in Figs. 5a and 7 for a given γ and R_0/b_0. However, in many biological systems, sometimes a few different types of protein cages co-exist in the same solution[42,43]. For instance self-assembly studies of dimeric Hepatitis B Virus capsid protein mutant Cp1492 shows that empty $T = 3$ and $T = 4$ structures form in a ratio of about 95:5 at medium to high salt concentration and close to neutral pH[44,45]. While the focus of our work is to find the optimal structure of protein cages as a function of mechanical properties of its building blocks, the polymorphism observed in several self-assembly studies can be explained through a careful examination of plots of energy per subunit versus number of subunits in Fig. 3. If the difference between the free energy per subunits in two different structures is small compared to the thermal energy $k_B T$, one expects to observe both structures, with relative populations given by the corresponding Boltzmann factor $\exp(\Delta \varepsilon / k_B T)$, with $\Delta \varepsilon$ the difference between the free energy per subunit in the two shells.

Lastly we emphasize that even though in this work we did not explicitly study the impact of salt and pH on the structure of protein shells, the solution environment such as salt and pH can modify the number of charges on the protein subunits, which in turn can change the stiffness and the spontaneous dihedral angle of building blocks. While the results obtained in this paper can explain why various protein shells with different symmetry appear in nature, at this point there is not enough experimental data to allow us to connect our variables R_0/b_0 and γ (the ratio of stretching to bending modulus) to the experimental conditions such as pH and salt.

To examine several concepts presented in this article, it would

be interesting to carry out a set of systematic experiments as a function of pH and salt concentration with various mutated proteins, which in consequence have different mechanical properties. One then can construct an experimental phase diagram similar to the one shown in Fig. 5b. A quantitative comparison between experiments and theory will result in a better understanding of the protein-protein interaction and the parameters that contribute to the formation of various protein cages with extensive potential application in various area of material science, gene delivery and medicine.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, *Nature*, 1985, **318**, 162.

2 R. H. Baughman, D. S. Galvão, C. Cui, Y. Wang and D. Tománek, *Chemical physics letters*, 1993, **204**, 8–14.

3 W. M. Aumiller, M. Uchida and T. Douglas, *Chemical Society Reviews*, 2018, **47**, 3433–3469.

4 J. Fontana, D. Nemecek, C. A. McHugh, A. A. Aksyuk, N. Cheng, D. C. Winkler, J. B. Heymann, E. Hoiczyk and A. C. Steven, *Microscopy and Microanalysis*, 2014, **20**, 1244–1245.

5 K. Ma, Y. Gong, T. Aubert, M. Z. Turker, T. Kao, P. C. Doerschuk and U. Wiesner, *Nature*, 2018, 1.

6 J. D. Perlmutter, C. Qiao and M. F. Hagan, *Elife*, 2013, **2**, year.

7 G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik and R. Zandi, *Physical Review E*, 2016, **94**, 022408.

8 J. G. Heddle, *Nanotechnology, science and applications*, 2008, **1**, 67.

9 R. F. Garmann, M. Comas-Garcia, C. M. Knobler and W. M. Gelbart, *Accounts of chemical research*, 2015, **49**, 48–55.

10 M. Chevreuil, D. Law-Hine, J. Chen, S. Bressanelli, S. Combet, D. Constantin, J. Degrouard, J. Möller, M. Zeghal and G. Tresset, *Nature communications*, 2018, **9**, 3071.

11 F. H. Crick and J. D. Watson, *Nature*, 1956, **177**, 473–5.

12 P. Weiss, *Science News*, 2005, **168**, 152–154.

13 J. E. Johnson and J. A. Speir, *Journal of molecular biology*, 1997, **269**, 665–675.

14 D. L. Caspar and A. Klug, *Cold Spring Harbor Symp. Quant. Biol.*, 1962, **27**, 1–24.

15 A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S. C. Harrison, T. Kirchhausen and T. Walz, *nature*, 2004, **432**, 573.

16 Y. Cheng, W. Boll, T. Kirchhausen, S. C. Harrison and T. Walz, *Journal of molecular biology*, 2007, **365**, 892–899.

17 R. Crowther, J. Pinch and B. Pearse, *Journal of molecular biology*, 1976, **103**, 785–798.

18 J. Min, S. Kim, J. Lee and S. Kang, *RSC Advances*, 2014, **4**, 48596–48600.

19 C. A. McHugh, J. Fontana, D. Nemecek, N. Cheng, A. A. Aksyuk, J. B. Heymann, D. C. Winkler, A. S. Lam, J. S. Wall, A. C. Steven and E. Hoiczyk, *The EMBO journal*, 2014, **33**, 1896–1911.

20 M. F. Hagan and R. Zandi, *Curr. Opin. Virol.*, 2016, **18**, 36–43.

21 T. Chen, Z. Zhang and S. C. Glotzer, *Langmuir*, 2007, **23**, 6598–6605.

22 T. Chen, Z. Zhang and S. C. Glotzer, *Proceedings of the National Academy of Sciences*, 2007, **104**, 717.

23 R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart and J. Rudnick, *Proceedings of the National Academy of Sciences*, 2004, **101**, 15556–15560.

24 S. Paquay, H. Kusumaatmaja, D. J. Wales, R. Zandi and P. van der Schoot, *Soft Matter*, 2016, **12**, 5708–5717.

25 J. Wagner and R. Zandi, *Biophysical journal*, 2015, **109**, 956–965.

26 T. Kohyama, D. Kroll and G. Gompper, *Physical Review E*, 2003, **68**, 061905.

27 G. M. Rotskoff and P. L. Geissler, *Proceedings of the National Academy of Sciences*, 2018, **115**, 6341–6346.

28 S. Fejer, D. Chakrabarti and D. Wales, *Nano Lett*, 2010, **4**, 219.

29 V. Sivanandam, D. Mathews, R. Garmann, G. Erdemci-Tandogan, R. Zandi and A. L. N. Rao, *Scientific Reports*, 2016, **6**, 26328.

30 S. D. Hicks and C. L. Henley, *Phys. Rev. E*, 2006, **74**, 031912.

31 A. Levandovsky and R. Zandi, *Physical review letters*, 2009, **102**, 198102.

32 Z. Li and H. A. Scheraga, *Proceedings of the National Academy of Sciences*, 1987, **84**, 6611–6615.

33 J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, 2nd edn., 2006.

34 T. Dokland, *Cellular and Molecular Life Sciences CMLS*, 1999, **56**, 580–603.

35 M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks III and V. S. Reddy, *Nucleic acids research*, 2008, **37**, D436–D442.

36 C. Hulo, E. De Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios and P. Le Mercier, *Nucleic acids research*, 2010, **39**, D576–D582.

37 J. Ning, G. Erdemci-Tandogan, E. L. Yufenyuy, J. Wagner, B. A. Himes, G. Zhao, C. Aiken, R. Zandi and P. Zhang, *Nature Communications*, 2016, **7**, 13689.

38 S. Li, P. Roy, A. Travesset and R. Zandi, *Proceedings of the National Academy of Sciences*, 2018.

39 T. Nguyen, R. Bruinsma and W. Gelbart, *Physical review letters*, 2006, **96**, 078102.

40 J. Lidmar, L. Mirny and D. R. Nelson, *Physical Review E*, 2003, **68**, 051910.

41 G. Vernizzi, R. Sknepnek and M. O. De La Cruz, *Proceedings of the National Academy of Sciences*, 2011, **108**, 4292–4296.

42 H.-K. Lin, P. van der Schoot and R. Zandi, *Phys. Biol.*, 2012, **9**, 066004.

43 R. Kusters, H.-K. Lin, R. Zandi, I. Tsvetkova, B. Dragnea and P. van der Schoot, *J. Phys. Chem. B*, 2015, **119**, 1869–1880.

44 P. Moerman, P. Van Der Schoot and W. Kegel, *The Journal of Physical Chemistry B*, 2016, **120**, 6003–6009.

45 P. T. Wingfield, S. J. Stahl, R. W. Williams and A. C. Steven, *Biochemistry*, 1995, **34**, 4919–4932.

Global minimum energy structures of protein nano-cages are investigated as a function of the mechanical properties of the coat proteins.
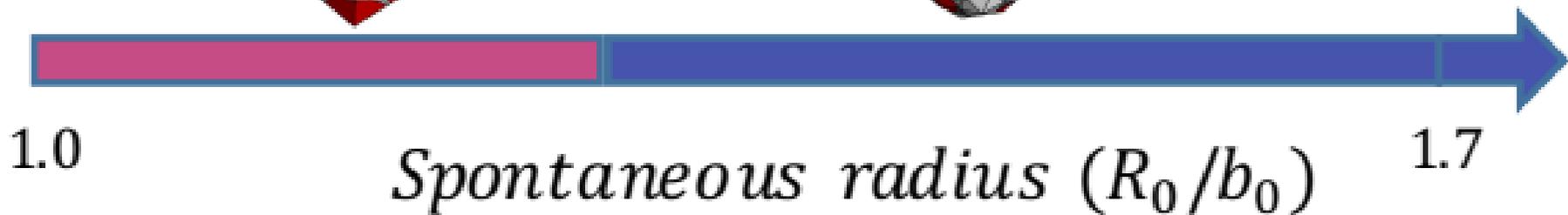
high γ (*rigid subunits*):

Nanoscale

1.0   *Spontaneous radius* (R_0/b_0)   1.7

low γ (*flexible subunits*):

1.0   *Spontaneous radius* (R_0/b_0)   1.7