The equilibrium structure of self-assembled protein nano-cages

Sanaz Panahandeh, Siyu Li, and Roya Zandi
Department of Physics and Astronomy, University of California, Riverside, California 92521, USA

Understanding how highly symmetric, robust, monodisperse protein 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 of building blocks including the spontaneous curvature, flexibility and bending rigidity of coat proteins are sufficient to predict the size of the assembly products and that the symmetry and shape selectivity of nano-cages can be explained, at least in part, on a thermodynamic basis. The polymorphism of nano-cages observed in vitro assembly experiments are also discussed.

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 plasmonic hydrocarbons, heat shock proteins, ferritins, carboxysomes, silicages, multicomponent ligand assemblies, clathrin vesicles and virus shells, to name a few [1, 5]. 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 addition to many other symmetric shells [15, 16] 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 [17], a Clathrin shell with tetrahedral symmetry [18], 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 nano-shells.

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 nano-shells 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.

Finally, we find that there are striking similarities between equilibrium and non-equilibrium “shape” phase diagrams as a function of the mechanical properties and spontaneous curvature of building blocks. The fact that pentamers form in the “correct” positions during the assembly process of symmetric shells, even under non-equilibrium conditions is quite unexpected. This reveals that there is a large affinity for the formation of disclinations (pentamers) at specific locations during the growth of most symmetric nano-cages, which are built from subunits that can form hexagonal sheets in flat space. Nevertheless, 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 non-equilibrium conditions.

It is 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]. The stretching energy of triangular network can simply be defined by a harmonic potential summed over all triangles,

\[ E_s = \sum_i \sum_{a=1}^3 \frac{k_s}{2} (b_{ia}^e - b_0)^2 \]  

with \( i \) the triangular subunit index, \( b_0 \) the equilibrium length of the edges, and \( b_{ia}^e \) the length of the \( a^{th} \) edge in the \( i^{th} \) 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,

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

with $<ij>$ the index pairs of neighboring subunits, $k_b$ the torsional spring constant, and $\theta_0$ the preferred dihedral angle between two subunits. The preferred dihedral angle and 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 $\theta_{ij}$ is between the unit normal vectors $\hat{n}_i$ and $\hat{n}_j$ of the two adjacent subunits $i$ and $j$ ($\cos \theta_{ij} = \hat{n}_i \cdot \hat{n}_j$), 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

$$\gamma = k_b b_0^2 / k_b,$$

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 \(^{31}\). We start from a triangulated spherical mesh with a random distribution of $N_v$ vertices. Each MC step consists of $N_v$ attempted edge swaps, which involves removing and reattaching the edge connecting two vertices of two neighboring triangles such that the two vertices which were not connected before, will be linked by an edge after the flip, as shown in Fig. 2. 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 \(^{32}\). The probability that the new relaxed structure with the new position of vertices to 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. Since our goal is to obtain the global minimum energy structure at low temperature, we relax the system after each edge swapping. 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 configurations many times. We also employ different cooling paths to avoid local minimum free energy traps for achieving equilibrium structures. Figure 2 illustrates a bond flip move. The thick edge in Fig. 2a is removed but then it will be linked to two other vertices in Fig. 2b. After that we allow the shell to relax as illustrated in Fig. 2c. 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 (Fig. 2).

This 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. Since in nature larger shells need some external help like scaffolding proteins or inner core to form symmetric shells \(^{33}\), the focus of this paper is on the smaller shells that are able to assemble spontaneously without any core.

**RESULTS**

We carry out Monte Carlo simulations as described in the previous section for different number of subunits $n_s$. We start with a fixed preferred spontaneous radius of curvature $R_0/b_0 = 1.28$ but different values of $\gamma$. The results of the simulations are illustrated in Fig. 3 in the form of a plot of the minimized elastic energy per triangles $\epsilon_n$ (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 once more that since $\gamma$ is proportional to the ratio of stretching to bending modulus, for larger $\gamma$s it is difficult to deform the subunits from their equilibrium equilateral shape but rather easy to bend them away from their preferred dihe-
dral angle. For small \( \gamma \)s, 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 structure for \( \gamma = 0.5 \). However as \( \gamma \) increases the local energy minima corresponding to certain structures become deeper and more pronounced and the \( T = 3 \) icosahedral shell becomes a global minimum for \( \gamma \geq 1 \). 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 \( \gamma \)s. Figure 4 illustrates the plot of energy per subunit versus \( R_0/b_0 \) for the global minimum energy structures (Fig. 3) at different \( \gamma \)-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. For instance, the letter \( A \) at the beginning of the dotted line shows that for \( \gamma = 8 \) the global minimum energy structure is a \( T = 1 \) icosahedral shell when \( 1 < R_0/b_0 < 1.3 \). Note that all the structures corresponding to the capital letters are illustrated in Fig. 5.

The 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 more apparent in Fig. 5, which is a plot of 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. 7 in the form of a “shape” phase diagram of spontaneous radius of curvature \( R_0/b_0 \) and \( \gamma \). Each shaded region in the diagram corresponds to a different shell whose structure and symmetry are illustrated in Fig. 5. Figure 7 shows that the structures become more sensitive to the spontaneous radius of curvature as \( \gamma \) 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. 8. 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 \( \gamma \)s 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 \( \gamma \)-values icosahedral structures are the minimum energy structures for the range of the spontaneous curvature studied, as illustrated in Fig. 7.

The largest shell obtained in Fig. 7 contains \( n_s = 80 \) triangles corresponding to a \( T = 4 \) structure for smaller \( \gamma \)-values. Note that at intermediate \( \gamma \)s, another equilibrium structure with the same number of subunits as \( T = 4 \) shell \( (n_s = 80) \) but different symmetry exists, which we label it as \( H^* \) in Figs. 5 and 7. 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 all previous studies. First of all, a review of literature shows that there are fewer \( T = 4 \) structure in nature \[34, 35\]. Furthermore, in Refs. \[22, 24\] only the \( D_{3h} \) 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 \[23\].

**Equilibrium versus non-equilibrium**

The structures obtained through the irreversible pathway are illustrated in the form of a phase diagram of the dimensionless ratio of bending to stretching modulus \( (\gamma) \) and the spontaneous radius of curvature in Fig. 9. As in the case of equilibrium phase diagram, each color refers to a different symmetric structure. While the shells in Fig. 9 grow following the local minimum free energy pathway, during the assembly process once a pentamer or hexamer forms, its position is permanently fixed. Thus the structures of assembled shells could be completely far from equilibrium.

We find it quite striking that the shell assembly along the local minimum free 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. 7 and 9. The shells that only appear in the non-equilibrium phase diagram are illustrated in Fig. 10. Two small regions (structures B and C) in the non-equilibrium phase diagram (Fig. 9) 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 shell, 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. 7 and 9.

The white area in the non-equilibrium 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 (44) with $D_2$ symmetry in that region. Further-
more, the regions corresponding to $G$ and $F$ structures cover a larger parameter space in the equilibrium phase diagram compared to the non-equilibrium one, revealing the presence of energy barriers preventing formation of some shells.

Furthermore, the irregular shells formed between the $G$ and $F$ structures at lower $\gamma$-values in the non-equilibrium 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 non-equilibrium one. In fact it only assembles at the boundary between the hashed and the white regions (Fig. 9), despite the fact that $K$ 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. 7 and 9 is $n_s = 80$. Note that in the absence of genome, an inner shell or scaffolding proteins, when we increase the spontaneous curvature at low $\gamma$ only irregular shapes form, and at large $\gamma$-values we obtain flat sheets or other structures with zero Gaussian curvature [30, 36, 37]. However, the focus of this paper is on the assembly of small symmetric shells, as illustrated in Figs. 7.

![Figure 8](image.png)

**FIG. 8.** Number of subunits in the equilibrium structures versus the spontaneous radius of curvature $R_0/b$ at $\gamma = 0.2$. The label next to each line shows the associated structure, illustrated in Fig. 6.

![Figure 10](image.png)

**FIG. 10.** The symmetric shells growing under non-equilibrium 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 non-equilibrium phase diagram (Fig. 9) they are labeled as B, C and E. The E structure has the same number of subunits as $E^*$ structure but with different symmetry.

**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 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 able to follow the general topological and geometrical constraints that all building blocks, no matter how complex or detailed, must obey. 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, i.e., they belong to the same “universality class” as coat proteins.

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. 7. 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 $\gamma$, where subunits can deform easily from the equilateral triangle, various structures form as a function of spontaneous curvature. However, by increasing $\gamma$, structures become more rigid and 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. [8]. However by 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. [7] was very similar to the phase diagram obtained under non-equilibrium conditions, see Fig. [9]. As explained in the previous section, the non-equilibrium 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. [7] and [9].

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]. The extensive similarities between equilibrium and non-equilibrium phase diagrams in the current paper indicate that during the growth process, for the symmetric shells other than icosahedral ones there are also deep potential wells for the formation of disclinations. While we cannot solve the continuum elasticity equations in the presence of spontaneous curvature and explain the symmetry of shells observed in Figs. [5] and [10] but based on our results, it is obvious that there is a delta-type potential 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 physical 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. [5] and [10] for a given $\gamma$ and $R_0/b_0$. However, in many biological systems, sometimes a few different types of protein cages co-exist in the same solution [39, 40]. For instance self-assembly studies of dimeric Hepatitis B Virus capsid protein mutant Cpl492 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 [41, 42]. 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 subunit in two different structures is small compared to the thermal energy $k_BT$, one expects to observe both structures, with relative populations given by the corresponding Boltzmann factor $\exp(\Delta\epsilon/k_BT)$, with $\Delta\epsilon$ 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 $\gamma$ (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. [7]. A quantitative comparison between experiments and the 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 application in various area of material science, gene delivery and medicine.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation through Grant No. DMR-1719550.

[1] H. W. Kroto, J. R. Heath, S. C. O’Brien, R. F. Curl, and R. E. Smalley, Nature 318, 162 (1985).
[2] R. H. Baughman, D. S. Galvão, C. Cui, Y. Wang, and D. Tománek, Chemical physics letters 204, 8 (1993).
[3] W. M. Auniller, M. Uchida, and T. Douglas, Chemical Society Reviews 47, 3433 (2018).
[4] J. Fontana, D. Nemecček, C. A. McHugh, A. A. Aksyuk, N. Cheng, D. C. Winkler, J. B. Heymann, E. Hoiczky, and A. C. Steven, Microscopy and Microanalysis 20, 1244 (2014).
