Statistical Physics of Viral Capsids with Broken Symmetry

L. E. Perotti,[1] J. Rudnick,[2] R. F. Bruinsma,[2,3,4] and W. S. Klug[1,4]
[1]Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, California 90095, USA
[2]Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA
[3]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA
[4]California NanoSystems Institute, University of California, Los Angeles, California 90095, USA
(Received 17 December 2014; published 29 July 2015)

We present a model to understand quantitatively the role of symmetry breaking in assembly of macromolecular aggregates in general, and the protein shells of viruses in particular. A simple dodecahedral lattice model with a quadrupolar order parameter allows us to demonstrate how symmetry breaking may reduce the probability of assembly errors and, consequently, enhance assembly efficiency. We show that the ground state is characterized by large-scale cooperative zero-energy modes. In analogy with other models, this suggests a general physical principle: the tendency of biological molecules to generate symmetric structures competes with the tendency to break symmetry in order to achieve specific functional goals.

DOI: 10.1103/PhysRevLett.115.058101 PACS numbers: 87.10.Hk, 05.10.−a, 87.16.aj

The protein shells, or capsids, that surround the genomes of viruses are frequently strikingly regular and elegant [1]. The capsids of most spherical viruses have the symmetry of one of the platonic solids: the icosahedron. In 1962 Caspar and Klug (CK) proposed a geometrical principle for the construction of icosahedral viral capsids [2]. It starts by drawing, on a hexagonal lattice, equilateral triangles whose vertices are located at the lattice sites, one of which is the origin. The triangles can then be indexed by the pair of integers h and k that determine the location of one of the two other vertices in terms of the two basis vectors of the hexagonal lattice. Icosahedra are then constructed by assembling together twenty such triangles, replacing a hexagon by a pentagon at each vertex. Different CK icosahedra are classified by their T *number* given by $h^2 + k^2 + hk = 1, 3, 4, 7, 9, 13, \ldots$. Figure 1 shows the case of $T = 9$ CK icosahedron with $(h = 3,\ k = 0)$. Actual capsids are formed by placing clusters of five (identical) proteins, or pentons, on the pentagons of a CK icosahedron and clusters of six proteins, or hexons, on the hexamers, adding up to a total of $60T$ proteins. The CK construction has remained a fundamental organizing principle of structural virology.

Proteins placed on the sites of a CK shell encounter T different local symmetries. Caspar and Klug argued that the CK construction maximizes the symmetry of the shell and hence minimizes the amount of intrinsic protein deformation generated by variations of the local environment. Given the success of the CK construction, it was surprising when detailed reconstructions of capsids by x-ray diffraction revealed numerous examples of capsids that had the symmetry of a CK shell but with capsid proteins at symmetry-inequivalent locations that had very different conformations [4,5]. In addition, some capsids violated the symmetry of the CK shell altogether [6]. These facts prompt a critical question for biophysics: how do we explain symmetry breaking in macromolecular assembly, and what function does it play?

Of particular relevance to this question are viral capsids that *switch* between locally symmetric and asymmetric conformations. A well-studied case is the $T = 7$ bacteriophage HK97 [7,8]. During initial assembly, HK97 hexons have a shear deformation of about 20 percent [9] that breaks local symmetry. When DNA is injected into the shell—as part of the virus assembly process—a capsid protein conformational change takes place as a result of which the hexons adopt a symmetric hexagonal shape, and the icosahedral capsid undergoes a *buckling* transition, with its morphology changing from spherical to faceted [10,11].

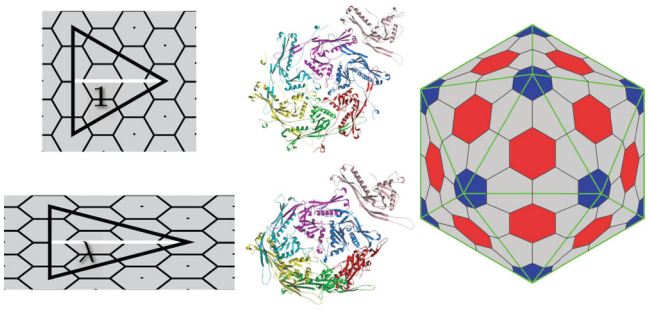

FIG. 1 (color online). Construction of a CK $T = 9$ icosahedron from an $(h = 3,\ k = 0)$ lattice vector creating an equilateral triangle. The perpendicular height of the triangle is taken to be 1. Blue: pentons. Gray: penton-bordering hexons. Red: face-centered hexons. Bottom left: lattice triangle constructed from an $(h = 3,\ k = 0)$ lattice vector of a stretched hexagonal lattice where the perpendicular height of the triangle is increased by a factor λ. Second column: example of hexamer protein configurations in the unstretched-mature (top) and stretched-immature (bottom) capsid. These examples refer to the Head II and Prohead II $T = 7$ HK97 virus (ViperDB [3]).

HK97 is not exceptional, but a member of a large class of viruses that probably share a common ancestry [12]: the eukaryote-infecting herpes viruses (*Herpesviridae*) and the prokaryote-infecting tailed DNA bacteriophages (*Caudovirales*). Their sizes range from $T = 4$ to $T = 52$. As far as is known, the capsid proteins of viruses belonging to this class share with HK97 the particular protein fold motif that generates the symmetry breaking. For certain T numbers, including $T = 7$, it is possible to construct a capsid from sheared hexons while retaining icosahedral symmetry, but for other T numbers, such as $T = 9$ and all T numbers for which T is divisible by 3 (see Supplemental Material I [13]), shearing is fundamentally incompatible with icosahedral symmetry because there are twenty "face-centered" hexons that have threefold symmetry axes passing through their center (Fig. 1). In fact, in the limit of large T numbers icosahedral symmetry *must* be broken in the minimum energy configuration since proteins located away from the fivefold sites must adopt the sheared minimum energy configuration of the flat protein sheet. In this Letter we show that the CK construction can be generalized to include hexon symmetry breaking. This generalized construction can be represented by a dodecahedral lattice model with a quadrupolar order parameter on each site. We will first discuss the general mathematical properties of this model and then use thin-shell elasticity theory to construct an explicit $T = 9$ CK shell assembled from sheared hexons to test it.

Assume that a CK icosahedral shell is assembled from a flat sheet of isosceles shear-stretched triangles, of the form shown in Fig. 1. Each triangle is constrained back into an equilateral shape by a "prestress" before they are fitted together to form an icosahedron. Every triangle can be in one of three states, depending on the direction of stretch with respect to the three pentagons that occupy the vertices of the triangle. After the CK shell has been assembled the constraints are released and the shell is allowed to relax to equilibrium—for a given set of orientational choices. The resulting state of the shell can be represented by a *dodecahedral* lattice model whose vertices coincide with the centers of the faces of the icosahedron. On each vertex, a double-headed arrow along the original stretch direction is placed, projecting along one of the three edges of the dodecahedron that meet at the vertex [Fig. 2(a)]. This construction generates 3^{20} states in total. Since the double-headed arrow represents a state of shear, it transforms under an icosahedral symmetry operation as a two-by-two symmetric tensor, so it will be referred to as a "quadrupolar" order parameter.

Each state α of the system can be codified by a set of 20 integers $\{n_i\}$, where $n_i = 1, 2, 3$ gives the shear direction with $i = 1, \ldots, 20$ a site index running over the dodecahedral lattice. Let $\mathcal{E}_\alpha$ be the energy of the state α. It should be invariant under any one of the 120 icosahedral symmetry operations.

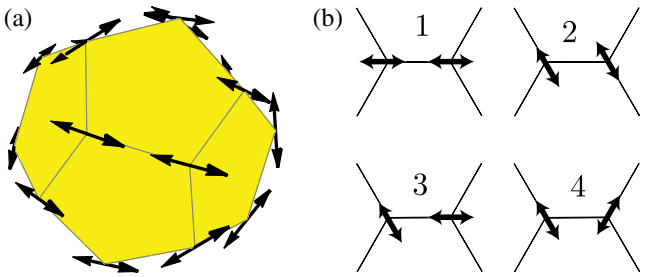

FIG. 2 (color online). Symbolic representation of an icosahedral CK shell constructed from stretched hexons. (a) The solid lines show the stretch direction at the centers of the twenty faces of the icosahedron, which span the dual dodecahedral lattice. The solid lines project onto one of the three edges of the dodecahedron that connect a site to its three neighbors. (b) Depiction of the four distinct bond patterns for pairs of neighboring order parameters.

By analogy with the lattice models of magnetism, it may or may not be possible to express the energy as a sum $\mathcal{E}_\alpha = \sum_{i,j} \epsilon(n_i, n_j)$ over pairwise interactions between neighboring quadrupoles with $\epsilon(n_i, n_j)$ a "bond" energy. For now, assume that this is the case. There are then $3 \times 3 = 9$ possible bond configurations for each pair of neighboring quadrupoles but only the four shown in Fig. 2(b) are not related by symmetry. Configurations 1 and 2 will be called "ferromagnetic" since their stretch directions are aligned (after parallel transport) while 3 and 4 will be called "antiferromagnetic" as they are more nearly perpendicular to each other. The corresponding bond energies are denoted by $\epsilon_{k=1,2,3,4}$. The bond occupation numbers $\mathcal{C}_{\alpha 1} - \mathcal{C}_{\alpha 4}$ are defined as the number of two-site configurations of each of the four bond species. The total (pairwise) interaction energy then equals

$$\mathcal{E}_\alpha = \mathcal{C}_{\alpha 1}\epsilon_1 + \mathcal{C}_{\alpha 2}\epsilon_2 + \mathcal{C}_{\alpha 3}\epsilon_3 + \mathcal{C}_{\alpha 4}\epsilon_4. \quad (1)$$

Because for every configuration α the total number of bonds $\sum_k \mathcal{C}_{\alpha k} = 30$ is fixed, changes in the bond occupation numbers must obey the condition

$$\Delta\mathcal{C}_1 + \Delta\mathcal{C}_2 + \Delta\mathcal{C}_3 + \Delta\mathcal{C}_4 = 0. \quad (2)$$

In Supplemental Material II [13], we provide a proof that there exists a second conservation law, which imposes

$$4\Delta\mathcal{C}_1 - 2\Delta\mathcal{C}_2 + \Delta\mathcal{C}_3 - 2\Delta\mathcal{C}_4 = 0. \quad (3)$$

These two conservation laws impose constraints on the bond energies. If the bond energies are shifted by $\epsilon_i \to \epsilon_i + \delta_i$ where $\delta_1 = \delta + 4\eta$, $\delta_2 = \delta - 2\eta$, $\delta_3 = \delta + \eta$, and $\delta_4 = \delta - 2\eta$ with δ and η arbitrary then the change in energy $\Delta\mathcal{E}_\alpha$ associated with any change in the bond occupation numbers is not affected by the shifts. This "gauge invariance" allows us to set the bond energies ϵ_2 and ϵ_3 equal to zero. From the two conservation laws, it

follows that a zero energy state should have ten type 2 bonds and twenty type 3 bonds. Whether zero energy states also are minimum energy states depends on the signs of ϵ_1 and ϵ_4.

In order to explore whether the pairwise decomposition assumption is valid and, if so, what the bond energies and minimum energy states are of the system, we numerically constructed a model capsid. It is known that certain physical properties of large capsids, such as the buckling transition between spherical and polyhedral capsid morphologies and the response to applied forces [14], are reproduced by thin-shell elasticity theory where the shell energy is written as the sum of a bending and a stretching term

$$\mathcal{E} = \frac{1}{2}\int dA\left[\kappa(2H)^2 + K(J-1)^2 + \mu\left(\frac{\mathrm{Tr}\overleftrightarrow{C}}{J} - 2\right)\right]. \quad (4)$$

Here, $\overleftrightarrow{C}$ is the right Cauchy-Green strain tensor. The first term is the bending energy, with κ the bending modulus and H the mean curvature. The second and third terms account for area dilation and isochoric shear, where K is the 2D (area) bulk modulus, $J = (\det\overleftrightarrow{C})^{1/2}$ is the deformed-to-reference area ratio, and μ is the 2D shear modulus. Energies will be expressed in units of κ and lengths in terms of the radius R of a spherical capsid. The two dimensionless numbers KR^2/κ and $\mu R^2/\kappa$ were set at 940 and 470, respectively. For these values, the Föppl–von Kármán number (FvK) $\gamma = (YR^2/\kappa)$, with Y the 2D Young's modulus, is equal to 1250, which is above the threshold for the buckling transition and similar to the value obtained for HK97 in its mature buckled morphology [10].

Before assembling the shell, shear-stretched hexons were constrained to adopt perfect sixfold symmetry so they could be fitted onto a $T = 9$ CK shell, after which the constraints were released. The energy was minimized following the method of Ref. [11] using both the steepest gradient and Monte Carlo methods. As a function of the stretching factor λ (see Fig. 1), a (reverse) buckling transition takes place with increasing λ from an icosahedral (outward pointing pentamers) to a dodecahedral (flat pentamers and outward pointing hexamers) morphology near $\lambda \simeq 1.06$, the point where the shell is spherical and the elastic energy has a minimum. A similar buckling transition from icosahedral to spherical is obtained by decreasing the capsid FvK number. However, changes in λ allow the additional spherical to dodecahedral transition and we focus here on the effect of different prestretch rather than its interplay with FvK number as already studied in Ref. [15]. Finally, we minimized the shell energy by allowing flips of the orientational configurations of the hexons. In all cases we found that in low energy states the stretch orientation of the sixty penton-bordering hexons was fixed along the hexon-penton interface while the orientations of the remaining twenty face-centered hexons fluctuated extensively. In order to obtain the excitation energies ϵ_j we rotated single hexons by steps of 120 degrees while maintaining all other states in the same orientation. As shown in Fig. 3, the ϵ_j can be obtained by computing changes in the deformation energy when a single hexon is rotated. If the pairwise approximation is valid, then the changes in energy should not depend on the background orientation of hexons that are not connected to the "test" hexon. Supplemental Material III [13] documents the evidence that the pairwise approximation accounts for the energy cost of order parameter flips with an error of about 6%. In Fig. 4, computed nonzero bond energies are plotted as a function of the stretch λ. Note the increase of ϵ_4 around the buckling transition where the shell transforms from icosahedral to dodecahedral.

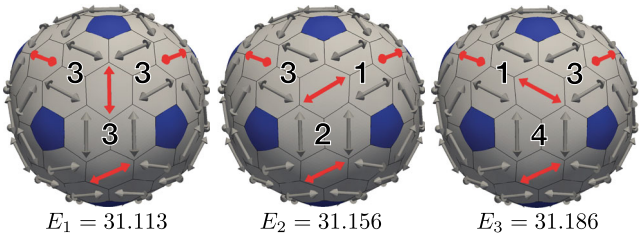

FIG. 3 (color online). Configurations used to compute energy tables: the orientation of the prestretch on the face-centered hexamer in the middle is rotated over 120 deg to compute the different types of interaction energies. The nonzero bond energies follow from the relations $E_2 - E_1 = \epsilon_1 = 0.043$ and $E_3 - E_2 = \epsilon_4 = 0.030$.

Since ϵ_1 and ϵ_4 are always positive, the zero energy states of the lattice model indeed are the minimum energy

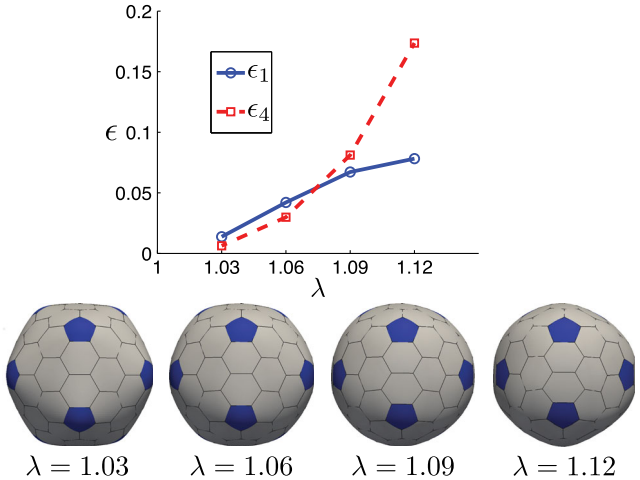

FIG. 4 (color online). Nonzero bond energies ϵ_1 and ϵ_4 in units of the bending energy κ as a function of the stretch λ. The corresponding capsid morphology is shown below the graph. Notice that, since the order parameter distribution is not icosahedral symmetric, it follows that the capsid morphologies are not strictly icosahedral or dodecahedral symmetric.

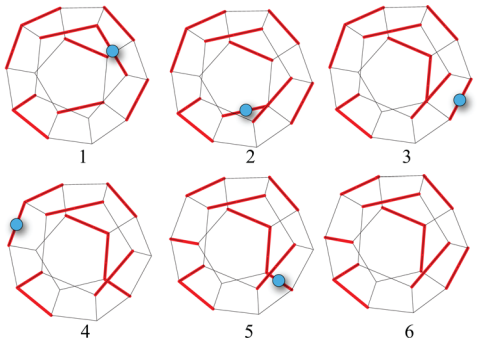

FIG. 5 (color online). A zero-energy mode: six zero energy states that can be transformed into each other by the rotation of the order parameter on one site. Type 2 bonds are shown thick and red, while type 3 are thin and black. The bond that is about to be moved is indicated by a dot.

states. Using the lattice model, we determined the energy spectrum for the bond energies $\epsilon_1 = 0.04$ and $\epsilon_4 = 0.03$ obtained for the buckling transition at $\lambda = 1.06$. When we randomly sampled 3×10^6 of the 3^{20} configurations we encountered only 16 zero energy configurations that were not related by an icosahedral symmetry operation while for 3×10^7 configurations, we encountered 155 such zero energy configurations. Based on this, we estimate that there are $\sim$2–3 $\times 10^2$ nonequivalent degenerate zero energy configurations. Figure 5 shows an example of six zero-energy configurations. The six states shown in Fig. 5 are part of a connected family of states that can be transformed into one other by successive single rotations, each one involving the relocation of one type 2 bond. Such a connected family of zero-energy states will be called a "zero mode." Typically, for every zero-energy state, there are multiple choices for zero-mode, single-spin rotations. The largest zero mode we found has 167 members, but there are also zero-energy states that are isolated in the sense that they are not connected to any other zero-energy state by a single rotation. Performing a systematic survey, we have found zero modes with 1, 4, 9, 11, 14, 41, 121, and 167 members.

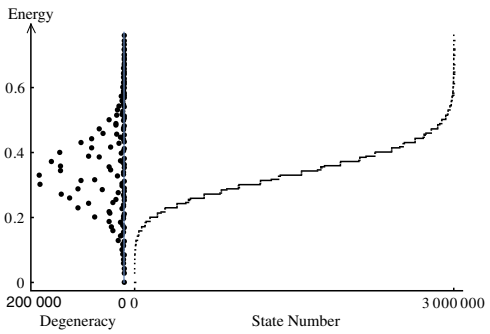

FIG. 6 (color online). Energy spectrum of the lattice model for $\epsilon_1 = 0.04$ and $\epsilon_4 = 0.03$ in units of the bending energy. Horizontal axis right: states ordered from low to high energy. Horizontal axis left: degeneracy level. Vertical axis: energy in units of the bending energy.

The spectrum of excited states for the 3×10^6 randomly sampled states is shown in Fig. 6. When states are rank ordered according to energy (right-hand side), it is seen that the spectrum has a funnel-like structure, reminiscent of the one encountered in the theory of protein folding [16]. The level degeneracy (left-hand side) increases dramatically as the funnel broadens out, resembling the molten-globule state of protein folding. The heat capacity has a pronounced maximum as a function of temperature when the thermal energy is comparable to the plateau energies of the spectral funnel (see Supplemental Material IV [13]).

We have shown that hexon symmetry breaking by shear stretching in the CK construction leads to a statistical mechanics model whose ground state is degenerate and characterized by cooperative zero-energy modes. These ground states lie at the bottom of a funnel of higher-energy states that have a rapidly increasing level of degeneracy. It is of interest to place the results in the context of the literature on the symmetry of protein oligomers. Functional protein oligomers formed from identical protein blocks often are symmetric [17], most likely because symmetric clusters allow for the formation of a maximum number of stable bonds [18]—an argument similar to that of CK. The tendency to generate symmetric protein structures with optimal thermodynamic stability and assembly kinetics competes with the tendency of biological molecules to *break* symmetry in order to achieve specific functional goals [17]. The breaking of icosahedral symmetry of the hexons of large viral capsids by shearing may be a functional feature of capsid assembly in this sense. Just as chaperonins promote protein folding by preventing unwanted interactions [16], hexon symmetry breaking may reduce the probability of capsid assembly errors by weakening bonding energies, hence allowing for more reversible, error-correcting assembly steps. Conformational changes following capsid assembly are then expected to increase the binding strength as hexon symmetry is restored, which is the case for HK97 [8,9].

We would like to acknowledge helpful discussions with R. Hendrix, J. Johnson, and A. Steven as well as support from the NSF under DMR Grants No. 1006128 and No. 1309423.

[1] T. Baker, N. Olson, and S. Fuller, Microbiol. Mol. Biol. Rev. **63**, 862 (1999).
[2] D. Caspar and A. Klug, in *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1962), Vol. 27, p. 1.
[3] M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks, and V. S. Reddy, Nucleic Acids Res. **37**, D436 (2009).

[4] M. G. Rossmann, Virology **134**, 1 (1984).
[5] J. E. Johnson and J. A. Speir, J. Mol. Biol. **269**, 665 (1997).
[6] J. M. Grimes, J. N. Burroughs, P. Gouet, J. M. Diprose, R. Malby, S. Ziéntara, P. P. Mertens, and D. I. Stuart, Nature (London) **395**, 470 (1998).
[7] J. Conway, W. Wikoff, N. Cheng, R. Duda, R. Hendrix, J. Johnson, and A. Steven, Science **292**, 744 (2001).
[8] A. Steven, B. Trus, F. Booy, N. Cheng, A. Zlotnick, J. Caston, and J. Conway, FASEB J. **11**, 733 (1997).
[9] I. Gertsman, L. Gan, M. Guttman, K. Lee, J. Speir, R. Duda, R. Hendrix, E. Komives, and J. Johnson, Nature (London) **458**, 646 (2009).
[10] J. Lidmar, L. Mirny, and D. R. Nelson, Phys. Rev. E **68**, 051910 (2003).
[11] A. Aggarwal, J. Rudnick, R. F. Bruinsma, and W. S. Klug, Phys. Rev. Lett. **109**, 148102 (2012).
[12] M. Baker, W. Jiang, F. Rixon, and W. Chiu, J. Virol. **79**, 14967 (2005).
[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.058101 for (i) the T number of capsids with face-centered hexagonal sites, (ii) the derivation of conservation law, (iii) the calculation of interaction energies, and (iv) calculation of the heat capacity.
[14] W. Roos, R. Bruinsma, and G. Wuite, Nat. Phys. **6**, 733 (2010).
[15] L. E. Perotti, A. Aggarwal, J. Rudnick, R. Bruinsma, and W. S. Klug, J. Mech. Phys. Solids **77**, 86 (2015).
[16] C. Dobson, *The Nature and Significance of Protein Folding* (Oxford University Press, Oxford, 2000).
[17] D. S. Goodsell and A. J. Olson, Annu. Rev. Biophys. Biomol. Struct. **29**, 105 (2000).
[18] J. Monod, in *Proceedings of the Nobel Symposium on Symmetry and Function of Biological Systems* (Wiley Interscience, New York, 1968), Vol. 11, p. 15.