Spontaneous curvature as a regulator of the size of virus capsids

Antonio Šiber[1,*] and Antonio Majdandžić[2]
[1]*Institute of Physics, Bijenička Cesta 46, 10000 Zagreb, Croatia*
[2]*Department of Physics, University of Zagreb, 10000 Zagreb, Croatia*
(Received 5 March 2009; revised manuscript received 29 June 2009; published 12 August 2009)

We investigate the physical reasons underlying the high monodispersity of empty virus capsids assembled in thermodynamical equilibrium in conditions of favorable *p*H and ionic strength. We propose that the high fidelity of the assembly results from the effective spontaneous curvature of the viral protein assemblies and the corresponding bending rigidity that penalizes curvatures which are larger and smaller from the spontaneous one. On the example of hepatitis B virus, which has been thoroughly studied experimentally in the context of interest to us, we estimate the magnitude of bending rigidity that is needed to suppress the appearance of aberrant capsid structures ($\sim 60 k_B T$). Our approach also demonstrates that the aberrant capsids that can be classified within the Caspar-Klug framework are in most circumstances likely to be smaller from the regular ones, in agreement with the experimental findings.

DOI: 10.1103/PhysRevE.80.021910 PACS number(s): 87.15.nr, 81.16.Dn, 82.60.Nh, 62.20.D−

I. INTRODUCTION

In physiological conditions, assembly of most viruses is precise which means that the percentage of aberrant structures is small [1,2]. This can, however, change depending on the *p*H value of the solution [3–5], presence or absence of specific ions [3], scaffolding proteins that assist the assembly of some viruses [6,7], and virus genetic material (DNA or RNA molecule) [8,9]. There are some viruses that assemble imprecisely i.e., produce observable amount of unfunctional and often irregular structures in the process of assembly. Perhaps the most notable of them are human immunodeficiency virions (HIV) which are very polydisperse [10]. One could argue that the presence of genome molecules, in particular their length, guarantees the precision of the assembly and formation of viruses of certain (functional) size. This issue has been recently investigated [11,12].

In this paper, we are interested in viruses that assemble spontaneously, i.e., without any specific molecular machinery that uses up chemical energy to form functional protein-genome complexes (this excludes, e.g., bacteriophages which use the adenosine triphosphaate (ATP) driven molecular motors to pack the DNA molecule in the preformed capsid [13,14]). The proteins that make a protective shell (capsid) of these viruses are known to self-assemble in (empty) capsids even in the *absence* of the genome molecule (mostly RNA for the viruses in question), at least when the salt concentration and *p*H value are appropriate [1,2,15,16] (these parameters obviously influence the protein-protein interactions). It is intriguing that the thus assembled capsids are often quite monodisperse and that their radii are the same as the radii of the functional viruses that contain the RNA molecule [2,15]. This suggests that there is some physical factor pertinent to protein-protein interactions in virus capsids that regulate the precision of the assembly. The deciphering of this factor is the motivation for this work.

According to the Caspar-Klug quasiequivalence theory of viruses [17], the same protein subunits may in principle assemble in many different capsids, each of which contains a certain prescribed number of protein subunits [protein subunits may be individual proteins, but also protein dimers which is the case e.g., in hepatitis B virus (HBV) [2]]. The total number of protein subunits (q) in viruses that obey the Caspar-Klug classification scheme is

$$q = 60\mathcal{T}, \tag{1}$$

where $\mathcal{T}$ is the triangulation (or quasiequivalence) order of a virus which can be expressed in terms of two integers h and k as

$$\mathcal{T} = h^2 + hk + k^2. \tag{2}$$

Capsids of different $\mathcal{T}$ numbers have different radii which are proportional to $\sqrt{\mathcal{T}}$. It has been shown that the electrostatics alone in combination with the simplest form of angle-independent hydrophobic attraction between the protein subunits cannot guarantee the formation of monodisperse empty capsids of certain preferred $\mathcal{T}$ number [18]. This may suggest that the protein-protein interactions are the strongest for a certain dihedral angle formed by neighboring protein subunits. In coarse-grained description, this translates in the existence of preferred (or spontaneous) curvature of viral capsids for which the capsid elastic bending energy is minimal so that the bending component of the elastic energy of the capsid can be written as

$$E_b = \frac{\kappa}{2}\int_S dS[c(\mathbf{r}) - c_0]^2. \tag{3}$$

Here, κ is the bending rigidity, dS is the infinitesimal element of the capsid surface (S), $c(\mathbf{r})$ is the mean curvature at point $\mathbf{r}$ on the capsid, and c_0 is the capsid spontaneous curvature. The value of the bending energy constant influences the total energy of the aberrant capsids so that it penalizes capsids with $\mathcal{T}$ numbers that are different (both larger and smaller) from the one pertinent to functional viruses. The question that arises concerns the required magnitude of κ such that the formation of aberrant capsids, both larger and smaller, in the equilibrium be suppressed, i.e., that the self-assembly is ac-

*asiber@ifs.hr

curate. The answer to this question is the main goal of this paper. There is a larger context of our work and we do think that our results may be of use to nanotechnology oriented research aiming to produce artificial systems that self-assemble reliably. Some recent work in this area seems to be promising, especially when it includes usage of viral proteins for production of nanocontainers [19–21].

In Sec. II we specify the model for the assembly that we use. In Sec. III we apply the model to the specific case of the assembly of hepatitis B capsids (empty viruses). The assembly of this virus has been thoroughly studied experimentally [1,2] and that is the reason that we choose it as a convenient benchmark for the application of our model. Section IV contains the discussion of the numerical results and their interpretation in the context of experimentally determined features of the virus assembly. Section V concludes the paper.

II. MODEL FOR THE ASSEMBLY OF REGULAR AND ABERRANT CAPSIDS IN THE EQUILIBRIUM

We shall model the assembly of capsids using the appropriately modified classical nucleation theory already developed in the context of viruses in Refs. [22,23]. Below, we briefly summarize the main features of the model of capsid nucleation when there is only a single path toward the assembled capsid. The density of the protein subunits in the solution is $\rho=N/V$, where N is the total number of subunits and V is the volume of the solution. In thermodynamical equilibrium, there are in general finite numbers of protein clusters (N_j) consisting of j units where $j=1$ represents the isolated subunits and $j=q$ the assembled capsids. The corresponding densities are denoted by ρ_j ($\rho_j=N_j/V$), and ρ_1 in particular is the density of the isolated subunits (monomers). For low protein concentrations, i.e., when $\rho\omega \ll 1$, the free-energy functional can be written (up to an additive constant that does not depend on N_j) as

$$\beta F = \sum_{j=1}^{q} [N_j \ln(\rho_j \omega) - N_j + \beta j \overline{\Delta g_j} N_j], \tag{4}$$

where ω is reference volume, interpreted as the free volume or cell volume in our approach—it should be thus of the order of magnitude that is characteristic for the water molecule. Although Eq. (4) was used with success in Ref. [23], its derivation and, especially, interpretation are by no means trivial. Perhaps the most troublesome of these problems is related to the determination of the exact value of the reference volume, ω. This is an old problem in cell models of the solutions that arises in part from the coarse grained description ("integration") of the solvent degrees of freedom. The characteristic scale is thus related to the mean volume of the solvent molecule. For details, see Ref. [24]. The exact value of ω shall in fact not be needed for our purposes.

The first two terms in the sum in Eq. (4) represent the entropy of j clusters, while the second term represents their binding free energy. The mean binding free energy per protein subunit is denoted by $\overline{\Delta g_j}$, where the overline emphasizes the fact that this quantity is assumed to be averaged over all subunits in the cluster (since the total binding free energy of the cluster is written simply as $j\overline{\Delta g_j}$, $\overline{\Delta g_j}<0$). The microscopic interpretation of this quantity is quite complex and includes additional averaging of protein-protein interactions over the cell volume.

Requiring that F be minimal in the equilibrium produces the set of equations that must be satisfied by the densities of protein clusters,

$$\rho_j \omega = \exp[-j(\beta \overline{\Delta g_j} + \mu)], \quad j = 1, \ldots, q, \tag{5}$$

where μ is the Largange multiplier associated with the requirement of conservation of the total number of subunits,

$$\rho = \sum_{j=1}^{q} j \rho_j. \tag{6}$$

Taking the reference state to be that of a single subunit in the solution ($\overline{\Delta g_1}=0$) allows us to write the equations for equilibrium densities as

$$\rho_j \omega = (\rho_1 \omega)^j \exp(-j\beta \overline{\Delta g_j}), \quad j = 2, \ldots, q, \tag{7}$$

which should again be solved on the hyperplane defined by the constraint in Eq. (6).

It is of some interest to present an alternative derivation of Eq. (7) that is quite similar to the standard theory of micellization (or polymerization) [25]. The system of dissolved proteins can be viewed as a mixture of phases of differently sized clusters. In equilibrium, the chemical potential per particle (protein) must be the same in all phases, otherwise the rearrangement will take place. Thus,

$$\begin{aligned} \mu &= \mu_1^0 + k_B T \ln(N_1/M) \\ &= \mu_2^0 + \frac{k_B T}{2} \ln(N_2/M) \\ &= \mu_3^0 + \frac{k_B T}{3} \ln(N_3/M) = \cdots, \end{aligned} \tag{8}$$

where factors 1/2,1/3,… take into account the fact that the chemical potential is to be calculated per protein. Quantities μ_1^0, μ_2^0,... represent the mean interaction free energy per protein. Equations (8) can also be written as

$$\frac{N_j}{M} = \left[\frac{N_1}{M}\right]^j \exp\left(-\frac{j\mu_j^0}{k_B T}\right), \tag{9}$$

which is again subject to constraint in Eq. (6). These equations are the same as Eq. (7) presuming the identification of μ_j^0 with Δg_j.

One can proceed and solve either Eqs. (5) and (6) (with $q+1$ unknowns) or Eqs. (7) and (6) (with q unknowns). This has already been done in Ref. [26] where it has been demonstrated that the probability of occurrence of intermediate protein clusters (unfinished capsids with $j=2,\ldots,q-1$) in equilibrium is quite small. However, in order to solve the equations for the densities of different clusters in equilibrium, one needs to appropriately model the dependence of the mean binding free energy, $\overline{\Delta g_j}$, on the number of subunits in a cluster, j. A simple and appealing procedure has been detailed in Ref. [23]. It was assumed there that each j cluster is a portion of a sphere of fixed radius R that misses a spheri-

cal cap whose area depends of the deviation of j from q, i.e., the number of subunits required by a particular cluster to close its spherical surface. Assuming that the interaction energies are short ranged so that they can be ascribed to subunit contacts, the binding free energy of a particular cluster can be written in terms of the free energies per subunit that is characteristic of a completed capsid, $\overline{\Delta g_q}$, and the line tension energy that reflects that fact that the subunits on the rim of the uncompleted sphere are differently coordinated from the subunits in a complete capsid. In particular, one easily finds that the number of subunits on a rim of a cluster with j subunits is given by

$$n_j^{rim} \approx \sqrt{4\pi}\sqrt{j\left(1-\frac{j}{q}\right)}. \tag{10}$$

Assuming that the binding energy per subunit on a rim is a certain fraction of the binding energy per subunit in the complete capsid, one obtains that

$$\overline{\Delta g_j} = \frac{\overline{\Delta g_q}}{j}[j - n_j^{rim}(1-f)], \tag{11}$$

where f is of the order of 0.5 and can be thought of as accounting for the reduced coordination of the subunits on a rim. It is convenient to choose f as

$$f = 1 - \frac{1}{2\sqrt{\pi(1-q^{-1})}} \tag{12}$$

since this produces $\overline{\Delta g_1}=0$ (when $q \gg 1$, f approaches a value of 0.718 which means that the subunits on a rim lose about a third of their neighbors). It will turn out that the details of energetics of the unfinished capsids are not of utmost importance, since it will be shown (in agreement with Ref. [26]) that in equilibrium only the subunits ($j=1$) and the completed capsids ($j=q$) are present in non-negligible amounts. A note on the assumption that the total energy can be approximately written as the sum of energy in protein contacts is in order here. The hydrophobic component of the protein-protein interaction is local and proportional to the area buried in protein contacts [22,25] so the assumption of its locality seems to be in order. The electrostatic component of interaction is in general long ranged, but it has been shown in Ref. [18] that it is strongly screened for viruses in physiological salt concentration so that it is proportional to the area of a capsid, i.e., to the number of subunits. The interplay of hydrophobic attraction and electrostatic repulsion has been demonstrated to dominate the energetics of capsid assembly [22], which vindicates the usage of Eq. (11).

We now proceed to construct the model that provides the equilibrium densities of various subunit clusters when there are *two* nucleation pathways—one leading to the capsid of the same radius and number of subunits (q_1) as in the functional virus and the other to the aberrant capsid consisting of q_2 subunits (Fig. 1). The two pathways separate at a certain cluster size ($k-1$) which is a precursor both to the formation of regular and aberrant capsids. We shall assume that $k=2$, i.e., that there are already two types of clusters containing only two subunits. In our formalism, these differ by the contact angle. We denote by $\rho_{j,1}$ ($\rho_{j,2}$) the density of regular

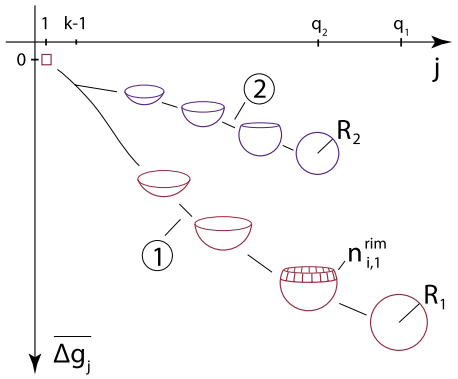

FIG. 1. (Color online) Schematic illustration of the pathways leading to (1) regular and (2) aberrant capsids. The quantities relevant for the model of assembly are denoted (see text).

(aberrant) j clusters. The binding free energies per subunit pertaining to protein cluster with l subunits on the regular pathway ($l=1,\ldots,q_1$) are denoted by $\overline{\Delta g_{l,1}}$. The analogous quantities for aberrant clusters are denoted by $\overline{\Delta g_{j,2}}$ ($j=2,\ldots,q_2$—note that we have positioned isolated subunits on the regular pathway which is merely a question of choice that does not influence the results).

We shall extend the model of assembly presented in Ref. [23] so to account for an alternative assembly pathway. The free-energy functional is now

$$\beta F = \sum_{j=1}^{q_1}[N_{j,1}\ln(\rho_{j,1}\omega) - N_{j,1} + \beta j\overline{\Delta g_{j,1}}N_{j,1}] + \sum_{j=k}^{q_2}[N_{j,2}\ln(\rho_{j,2}\omega) - N_{j,2} + \beta j\overline{\Delta g_{j,2}}N_{j,2}], \tag{13}$$

where $N_{j,1}$ ($N_{j,2}$) is number of regular (aberrant) clusters of size j and $\rho_{j,1}$ ($\rho_{j,2}$) is their density. This leads to the following set of equations for the densities of protein clusters in equilibrium:

$$\rho_{j,1}\omega = (\rho_1\omega)^j \exp(-j\beta\overline{\Delta g_{j,1}}), \quad j=2,\ldots,q_1$$

$$\rho_{j,2}\omega = (\rho_1\omega)^j \exp(-j\beta\overline{\Delta g_{j,2}}), \quad j=k,\ldots,q_2$$

$$\rho = \sum_{j=1}^{q_1} j\rho_{j,1} + \sum_{j=k}^{q_2} j\rho_{j,2}. \tag{14}$$

To solve these equations, a model for the binding free energies of aberrant clusters is required. We assume that the *only* difference in energetics is due to an unfavorable curvature of the aberrant capsids. For complete spherical capsids, Eq. (3) simplifies to

$$\frac{E_b}{4\pi R^2} = \frac{1}{A}\frac{E_b}{q} = \frac{\kappa(c-c_0)^2}{2}, \tag{15}$$

where R is the capsid radius, $c=1/R$, and A is the mean area per protein subunit in a completed capsid consisting of q subunits. Note that E_b/q is the excess bending energy per subunit. Since the incomplete clusters are assumed to be spheres with missing caps, the same relation holds also for

the excess bending energy per subunit in an incomplete cluster (the same holds for all portions of the sphere that have the same area). For the binding free energies per subunit we can thus write

$$\overline{\Delta g_{i,1}} = \frac{\overline{\Delta g_{q_1,1}}}{i}[i - n_{i,1}^{rim}(1 - f_1)],$$

$$\overline{\Delta g_{j,2}} = \frac{\overline{\Delta g_{q_1,1}}}{j}[j - n_{j,2}^{rim}(1 - f_2)] + A\frac{\kappa(c - c_0)^2}{2}, \quad (16)$$

where $i=1,\ldots,q_1$, $j=2,\ldots,q_2$, and c (c_0) now stands for the spontaneous mean curvature of the aberrant (regular) capsid. The f factors are given as in Eq. (12) with q replaced by q_1 and q_2 for the regular and aberrant subunits, respectively. The number of subunits on the rim is given as

$$n_{i,1}^{rim} = \sqrt{4\pi}\sqrt{i\left(1 - \frac{i}{q_1}\right)},$$

$$n_{j,2}^{rim} = \sqrt{4\pi}\sqrt{j\left(1 - \frac{j}{q_2}\right)}. \quad (17)$$

Note again that our modeling of the energetics of capsids reflects out assumption that the sole difference in binding energy in complete aberrant and regular capsids comes from the bending energy. Note that the number of subunit contacts does not depend on the capsid T number—the total number of contacts is $90T$, which means that there are 3/2 contacts per subunit on the average, irrespective of the capsid T number.

The model now contains only two, in general, unknown quantities: the binding free energy per subunit in a complete capsid ($\overline{\Delta g_{q_1,1}}$) and the bending rigidity (κ). The density of subunits in a completely disassembled state (ρ) is a parameter of the model, while A, q_1, and q_2 can be determined from the structure and size of the regular and aberrant version of the virus in question.

III. APPLICATION OF THE MODEL TO THE ASSEMBLY OF HEPATITIS B CAPSIDS

A. Full model of the assembly

We shall now apply the model elaborated in the previous section to the case of capsid of HBV. It was shown in Ref. [1] that the HBV proteins lacking 34 amino acids at the C terminal of the full-length capsid protein (183 amino acids) assemble in two different empty capsids—one that has 120 protein dimers, $Cp149_2$, ($\mathcal{T}$="4" structure, $q_1=120$) and the other, smaller, composed of 90 dimers ($\mathcal{T}=3$, $q_2=90$). The percentage of $\mathcal{T}=3$ capsids was found to be about 5%. When the full-length capsid proteins were further truncated, the percentage of $\mathcal{T}=3$ capsids grew, reaching ~85% for 140-residue protein. An extensive study regarding thermodynamics of assembly of empty capsids composed of $Cp149_2$ subunits was performed in Ref. [2]. This study enables us to extract the parameters relevant to our modeling. In particular, it was found that the $Cp149_2$-$Cp149_2$ contact energy in complete $\mathcal{T}$="4" capsids is on the order of 150 meV ($5.8k_BT$), depending both on temperature and salt concentration [2]. This allows us to fix one of the parameters of our model and here we choose a value that corresponds to the one derived in Ref. [1] for viruses in 0.3 M NaCl solutions at 25 °C. This yields $\overline{\Delta g_{q_1,1}}=322$ meV ($12.5k_BT$; 240 contacts were assumed in Ref. [1] for a capsid consisting of 120 subunits). A word of caution is in order here since the subunit of hepatitis B virus is a protein *dimer*—in our terminology *dimeric* Cp149 protein is to be treated as the basic subunit of the assembly, i.e., as a *monomeric* cluster (dimer in our terminology thus contains two Cp149 dimers, i.e., four Cp149 proteins in total).

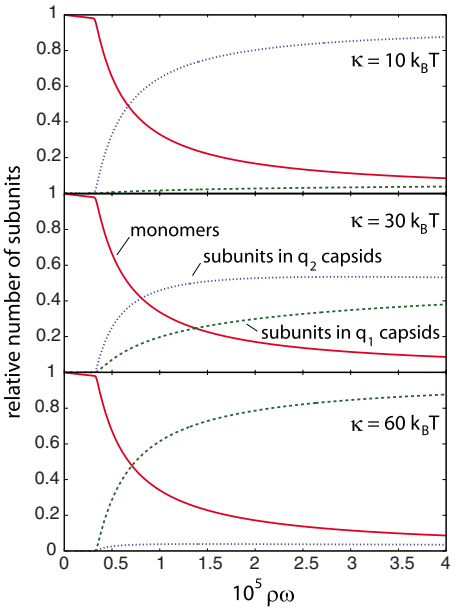

FIG. 2. (Color online) Relative number of $Cp149_2$ subunits as "monomers" (N_1/N—solid red line), assembled in regular capsids ($q_1N_{q_1}/N$—dashed green line), and in aberrant capsids ($q_2N_{q_2}/N$—dotted blue line).

The bending rigidity of virus capsids has been considered before and depending on a model used and the virus studied, different (but not too different) values have been suggested. For example, in Ref. [27], a value of $\kappa=(10-15)k_BT$ has been proposed, whereas value of $\kappa=40k_BT$ has been proposed as the upper limit for bending rigidity of HK97 bacteriophage. In the following, we shall treat κ as unknown parameter and examine how it influences the presence of $\mathcal{T}=3$ capsids in the equilibrium. This information is shown in Fig. 2. In these calculations, we have taken $A=22.5$ nm^2 which was calculated from the mean radius of $\mathcal{T}$="4" capsid (14.7 nm), and the preferred curvature was set to be equal to inverse mean radius of the capsid ($c_0=0.068$ nm^{-1}). The curvature of the $\mathcal{T}=3$ capsid is $c=c_0\sqrt{4}/\sqrt{3}=0.0785$ nm^{-1}. Relative numbers of $Cp149_2$ subunits that remain as "monomers," and assemble in regular and aberrant capsids are shown as functions of the appropriately scaled total volume ratio of dissolved subunits.

A feature of our results that is also present in the previous studies [23,26] is appearance of (pseudo)critical concentration of the protein subunits. Below the critical concentration,

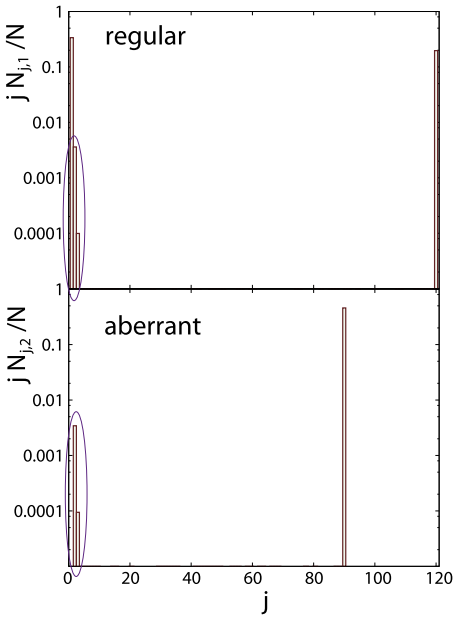

FIG. 3. (Color online) Relative number of $Cp149_2$ subunits in variously sized regular and aberrant clusters. Note that the single subunits belong to the regular pathway as explained in the text. The volume ratio chosen in this calculation is $\rho\omega=10^{-5}$, and the bending rigidity is $\kappa=30k_BT$.

$\rho\omega \sim 3.5\times10^{-6}$ in our calculations, the assembly does not take place and most of the subunits remain as monomers in the solution. It is not quite easy to observe that the sum of relative amounts of subunits in monomers, regular, and aberrant capsids is not equal to one. This is because the calculations predict that in equilibrium a very small percentage of subunits assembles in other forms, i.e., clusters or unfinished capsids. This is shown in Fig. 3 where one can clearly see nonvanishing concentrations of clusters containing two and three $Cp149_2$ subunits, but note that the total relative amount of dimers (both regular and aberrant) is smaller than 1%. The concentrations of all the j clusters are nonvanishing but their numerical values are extremely small.

We shall now briefly pause to interpret the obtained results. In the thermodynamical equilibrium, the system settles in a minimum of free energy. For finite temperatures, this is a state of compromise between the internal energy being as low as possible and entropy being as high as possible. This immediately suggests that aberrant capsids, *larger* than regular ones, are unfavorable both in terms of energy and entropy so they should not exist in equilibrium. Indeed, that is what we have verified in our numerical calculations by performing simulations with $q_2>q_1$ (not shown; the amount of complete aberrant capsids larger than regular ones is vanishingly small even for $\kappa=0$)—this argument, of course, heavily depends on the assumption that the curvature of the regular capsids *is the same* as the spontaneous curvature. On the other hand, aberrant capsids which are *smaller* than regular ones, although unfavorable in terms of energy (because of the bending contribution to the capsid energy), are favorable in terms of entropy (due to the fact that they are smaller, the kinetic or mixing part of the translational entropy is favorable). There is thus a subtle interplay between the energetics of the capsid and the entropy of their ensemble that is tuned by the value of κ which determines the probability of occurrence of aberrant capsids, as is clear from Fig. 2.

The scarcity of the intermediate structures in the equilibrium *a posteriori* corroborates the usage of our simple model which takes into account only two pathways. One could construct a mathematical procedure so to keep track of different pathways going to intermediates of different symmetry, i.e., one could account for the fact that there may be several geometrical realization of an intermediate cluster of size j. This has been done in Ref. [28]. However, an extremely small probability of occurrence of intermediates that we obtain in our approach suggests that such refined approach would not change the results significantly. For example, for the case shown in Fig. 3, the relative amount of proteins in $j=15$ regular clusters is 3.4×10^{-15} and in $j=119$ clusters (i.e., regular capsids missing one subunit) only 8.9×10^{-7}. In fact, an excellent agreement with our numerical data could be obtained by immediately dropping all the intermediate clusters from the free energy, keeping only the monomers and the two types of capsids. This is a model that has been applied for regular capsids only in Ref. [23]. We now specify it for the case of our interest.

B. Minimal numerical model for the assembly

Setting the concentrations of intermediate clusters in equilibrium to zero, one obtains a single transcendental equation that connects the relevant quantities for the assembly,

$$\rho\omega = \rho_1\omega + q_1(\rho_1\omega)^{q_1}\exp(-q_1\beta\overline{\Delta g_{q_1,1}}) + q_2(\rho_1\omega)^{q_2}\exp(-q_2\beta\overline{\Delta g_{q_2,2}}), \tag{18}$$

where ρ_1 is, as before, the density of monomers in the equilibrium. Note that the modeling of the energetics of intermediate clusters is not needed now, and only the spontaneous curvature—bending energy term persists as the (unknown) parameter of the model (in addition to the binding energy per protein subunit). The number of parameters is thus formally the same as in the full model, but the details of the energetics of the intermediates that had to be accounted for by the full model (such as parameters f and n_j^{rim}) are of no relevance now. Unfortunately, Eq. (18) is still not amenable to analytic solutions and one again has to resort to numerical procedures in order to solve it. Nevertheless, its appeal lies in its simplicity and the small number of parameters that figure in it. Its solutions for the same sets of parameters used to produce Fig. 2 is shown in Fig. 4. Note that the numerical results are almost indistinguishable from those obtained assuming a finite concentration of intermediates in the thermodynamical equilibrium (full model). The clearest differences can in fact be seen for protein concentrations smaller from the critical one. For the full assembly model, in this region one obtains a slight (but visible) drop in the number of isolated protein subunits (monomers) due to the fact that small but not entirely negligible number of dimers form (see Fig. 3). Since dimers and all larger clusters are not included in the minimal model, the number of monomers remains practically one (1) for subunit concentrations smaller than the critical one.

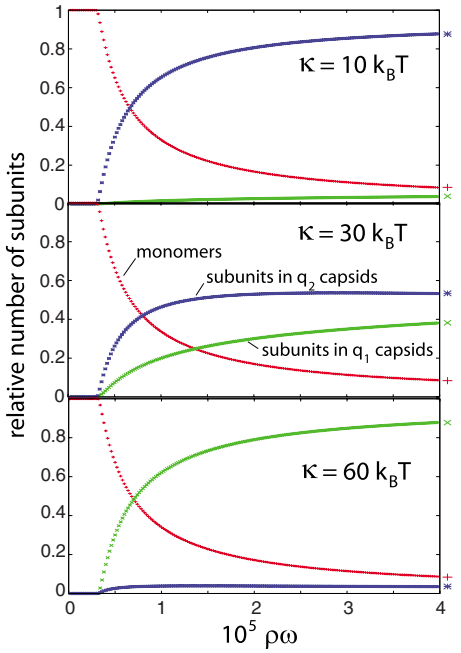

FIG. 4. (Color online) Relative number of $Cp149_2$ subunits as monomers (N_1/N—red plus symbols), assembled in regular capsids ($q_1 N_{q_1}/N$—green x symbols), and in aberrant capsids ($q_2 N_{q_2}/N$—blue stars). The results were obtained by solving the minimal numerical model for the assembly [Eq. (18)].

IV. DISCUSSION

It is interesting to note that the contribution of bending to the total binding energy of the protein contact is quite small. For $\kappa=60k_BT$, this contribution is only $0.07k_BT$ per contact in the aberrant $\mathcal{T}=3$ capsid, which is only 0.6% of the binding energy per protein subunit. Somewhat surprisingly, quite small contribution to the binding energy of the protein contacts is sufficient to effectively suppress the appearance of smaller aberrant structures in the equilibrium.

One could ask whether the change in pH and salinity of the solution influences the magnitude of the bending rigidity, κ. As κ is a coarse-grained parameter stemming from the microscopic details of protein-protein interactions, which are influenced by pH and salinity, the affirmative answer to this question seems to be in order. This suggests that the appearance of aberrant structures may be stimulated by changes in pH and salinity (ionic strength). Indeed, it has been found that the dominant structures observed in the solution of viral proteins strongly depend on these parameters (see, e.g., a recent study in Ref. [29]). One should, however, keep in mind that pH value and ionic strength influence directly the electrostatic interaction between the protein subunits which may in part be responsible for the bending rigidity of the capsid but may also contribute to the much larger angle-independent binding energy of the protein contacts.

The aberrant structures may not always easily fit in the Caspar-Klug classification scheme. Cylindrical assemblies, multishelled assemblies possibly consisting of concentric shells of different $\mathcal{T}$ numbers, hexagonal protein sheets, and isolated subunits are often found in the phase diagram of the virus assembly. Conical structures of different angles are common in the assembly of HIV proteins without a genome [10]. The part of the virus assembly phase diagram that consists of isolated subunits is most easily explained. Namely, the changes in pH and ionic strength may decrease the interaction between the protein subunits so to drive the solution below the density threshold needed for the onset of assembly. This has also been experimentally confirmed in the experiments by Ceres and Zlotnick [2]. In addition to irregular and non-Caspar-Klug structures, aberrant structures that can be classified within the Caspar-Klug framework are also typical in the assembly of virus capsids and have been found, in addition to hepatitis B virus, in cowpea chlorotic mottle virus [15] (regular: $\mathcal{T}=3$, aberrant: $\mathcal{T}=$"2"), P22 bacteriophage [6] (regular: $\mathcal{T}=7$, aberrant: $\mathcal{T}=4$), P2 bacteriophage [13] (regular: $\mathcal{T}=7$, aberrant: $\mathcal{T}=4$), and brome mosaic virus in the presence of RNA (regular: $\mathcal{T}=3$, aberrant: $\mathcal{T}=$"2"). The scarcity of these structures *in vivo* conditions may point to effective bending rigidity in such conditions being above threshold for the appearance of aberrant structures. Note again here that the aberrant structures are always smaller from the regular ones which can be easily explained using the thermodynamical arguments we presented earlier.

Scaffolding proteins are known to mediate the assembly of capsids in some viruses, bacteriophages in particular [13]. In the absence of these proteins, the capsid proteins are known to assemble in aberrant structures with larger probability. A nice example can be found in bacteriophage P22 procapsid whose proteins in the absence of scaffolding proteins assemble in $\mathcal{T}=7$ (regular) and $\mathcal{T}=4$ capsids [6]. Note again that the aberrant capsid is smaller from the regular one. It is of interest to note here that the scaffolding proteins may effectively act to finely tune the bending rigidity of the capsid protein assemblies so that the regular structure is assembled with large certainty. Special mechanisms to guarantee the formation of the regular capsids may be more crucial for the formation of large $\mathcal{T}$ number capsids. Namely, in this case, there is always a possibility to form capsids of smaller $\mathcal{T}$ numbers which are entropically favored so specific molecular machinery may be needed to suppress these effects. Larger viruses can contain longer genomes so that they can code for special chaperoning proteins that may be required in the process of correct assembly of large structures. A spectacular effect related to aberration of the virus capsids can be found in bacteriophage P2 and the assembly parasite P4. Namely, when the cell infected with P2 is superinfected with P4, the P2 capsid proteins do not assemble in $\mathcal{T}=7$ structure, but only in smaller $\mathcal{T}=4$ structure which is too small to host the P2 genome, but large enough to host the P4 genome (since P4 genome does not contain the code for capsid proteins). Again, the interplay of many proteins involved in the assembly of these viruses may act to effectively change the spontaneous curvature of the protein assemblies thus forming entropically preferred smaller capsids.

An important question that has not been answered by our analysis is the origin of the spontaneous curvature term and the bending rigidity. A particularly appealing microscopic interpretation of these parameters concerns the role of attractive hydrophobic interactions and their dependence on the dihedral angle of protein contacts. Namely, one could argue that the buried protein surface is the largest for the capsid of

regular $\mathcal{T}$ number and that it becomes smaller for larger and smaller aberrant capsids. This would then translate in effective weakening of the protein-protein interactions that may be described using the coarse-grained concept of spontaneous curvature. It is interesting to note in this context that the proportion of aberrant HepB capsids could be varied in experiments [1] by truncating the C terminus of the capsid protein. Within our theoretical framework, this procedure would result in the change in bending rigidity of the protein assemblies and/or change in the spontaneous curvature. In this way we could explain the appearance of $\mathcal{T}=3$ structures as the dominant ones for sufficiently truncated capsid proteins.

V. CONCLUSIONS

We have assumed that there is a spontaneous curvature of the virus capsids, resulting from the preferred angle of contacts between their protein subunits. From a numerical analysis of a relatively simple model of the self-assembly, we have demonstrated that the bending energy constant of the virus capsids acts as a kind of a monodispersity parameter, allowing or blocking the appearance of aberrant capsids in the thermodynamical equilibrium. We have found that there is a threshold magnitude of the bending rigidity that is needed in order to suppress the appearance of aberrant capsids that have $\mathcal{T}$ numbers smaller than the regular ones. For the case of hepatitis B virus, we find that the threshold value of κ is of the order of $60k_BT$—all the values of κ larger than this one will result in assemblies that are more monodisperse than those observed experimentally (one needs to keep in mind, however, that the experimental studies were done with the truncated versions of HepB capsid proteins). To suppress the appearance of aberrant capsids with $\mathcal{T}$ numbers larger than the regular ones, the bending rigidity is not essential as mixing entropy favors the appearance of smaller subunit assemblies.

In the end, it is pleasing to see that the estimates of bending rigidity obtained thus far for different viruses and using quite different methods [27,30] all fall in the range of several tens of k_BTs.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science, Education, and Sports of Republic of Croatia (Project No. 035-0352828-2837).

[1] A. Zlotnick, N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, S. J. Stahl, and P. T. Wingfield, Biochemistry **35**, 7412 (1996).
[2] P. Ceres and A. Zlotnick, Biochemistry **41**, 11525 (2002).
[3] D. M. Salunke, D. L. D. Caspar, and R. L. Garcea, Biophys. J. **56**, 887 (1989).
[4] L. S. Ehrlich, T. Liu, S. Scarlata, B. Chu, and C. A. Carter, Biophys. J. **81**, 586 (2001).
[5] M. Cuillel, C. Berthet-Colominas, P. A. Timmins, and M. Zulauf, Eur. Biophys. J. **15**, 169 (1987).
[6] P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King, and W. Chiu, Biophys. J. **74**, 559 (1998).
[7] P. E. Prevelige Jr., D. Thomas, and J. King, Biophys. J. **64**, 824 (1993).
[8] J. M. Johnson, D. A. Willits, M. J. Young, and A. Zlotnick, J. Mol. Biol. **335**, 455 (2004).
[9] M. A. Krol, N. H. Olson, J. Tate, J. E. Johnson, T. S. Baker, and P. Ahlquist, Proc. Natl. Acad. Sci. U.S.A. **96**, 13650 (1999).
[10] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, Science **283**, 80 (1999).
[11] A. Šiber and R. Podgornik, Phys. Rev. E **78**, 051915 (2008).
[12] R. Zandi and P. van der Schoot, Biophys. J. **96**, 9 (2009).
[13] T. S. Baker, N. H. Olson, and S. D. Fuller, Microbiol. Mol. Biol. Rev. **63**, 862 (1999).
[14] W. M. Gelbart and C. M. Knobler, Phys. Today **61**(1), 42 (2008).
[15] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, Virology **277**, 450 (2000).
[16] C. Berthet-Colominas, M. Cuillel, M. H. J. Koch, P. Vachette, and B. Jacrot, Eur. Biophys. J. **15**, 159 (1987).
[17] D. L. D. Caspar and A. Klug, Cold Spring Harb Symp. Quant. Biol. **27**, 1 (1962).
[18] A. Šiber and R. Podgornik, Phys. Rev. E **76**, 061906 (2007).
[19] C. Chen, M.-C. Daniel, Z. T. Quinkert, M. De, B. Stein, V. D. Bowman, P. R. Chipman, V. M. Rotello, C. Cheng Kao, and B. Dragnea, Nano Lett. **6**, 611 (2006).
[20] J. Sun, C. DuFort, M.-C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. Cheng Kao, and B. Dragnea, Proc. Natl. Acad. Sci. U.S.A. **104**, 1354 (2007).
[21] M. Uchida, M. T. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. O. Liepold, M. Young, and T. Douglas, Adv. Mater. **19**, 1025 (2007).
[22] W. K. Kegel and P. van der Schoot, Biophys. J. **86**, 3905 (2004).
[23] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, Biophys. J. **90**, 1939 (2006).
[24] H. Reiss, W. K. Kegel, and J. Groenewold, Ber. Bunsenges. Phys. Chem **100**, 279 (1996); H. Reiss and W. K. Kegel, J. Phys. Chem. **100**, 10428 (1996); K. Tokiwano and K. Arakawa, Bull. Chem. Soc. Jpn. **50**, 2217 (1977).
[25] J. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, New York, 1991).
[26] A. Zlotnick, J. Mol. Biol. **241**, 59 (1994).
[27] T. T. Nguyen, R. F. Bruinsma, and W. M. Gelbart, Phys. Rev. E **72**, 051923 (2005).
[28] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, Protein Sci. **14**, 1518 (2005).
[29] L. Lavelle, M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, and J. Ruiz-Garcia, J. Phys. Chem. B **113**, 3813 (2009).
[30] A. Šiber, Phys. Rev. E **73**, 061915 (2006).