

# Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses

Antonio Šiber<sup>1,\*</sup> and Rudolf Podgornik<sup>2,3</sup>

<sup>1</sup>*Institute of Physics, Bijenička Cesta 46, 10000 Zagreb, Croatia*

<sup>2</sup>*Department of Theoretical Physics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia*

<sup>3</sup>*Department of Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia*

(Received 24 July 2008; published 18 November 2008)

We investigate the influence of salt concentration, charge on viral proteins and the length of single-stranded RNA (ssRNA) molecule on the spontaneous assembly of viruses. Only the nonspecific interactions are assumed to guide the assembly, i.e., we exclude any chemical specificity that may lock the viral proteins and ssRNA in preferred configurations. We demonstrate that the electrostatic interactions screened by the salt in the solution impose strong limits on viral composition that can be achieved by spontaneous assembly. In particular, we show that viruses whose ssRNA carries more than twice the amount of charge that is located on the viral proteins, cannot be assembled spontaneously. We find that the spatial distribution of protein charge is important for the energetics of the assembly. We also show that the pressures that act on the viruses as a result of attractive protein-ssRNA electrostatic interactions are at least an order of magnitude smaller than is the case with bacteriophage viruses that contain double-stranded DNA molecule.

DOI: [10.1103/PhysRevE.78.051915](https://doi.org/10.1103/PhysRevE.78.051915)

PACS number(s): 87.15.nr, 41.20.Cv, 82.35.Rs

## I. INTRODUCTION

The study of shape and composition of viruses is a field of long tradition and rich activity. That there is some interesting mathematics and physics involved in this field has been well known, at least since the time when Donald Caspar and Aaron Klug published their paper on the physical principles of construction for the then-called “regular” viruses [1]. They have convincingly demonstrated that viruses can be thought of as triangulations of a sphere with an icosahedral “backbone.” This mathematical principle predicts that all proteins that make the coating of a virus (capsid) are not in identical surroundings concerning the positions and orientations of their neighboring proteins. However, of all possible platonic triangulation schemes and hypothetical “viruses” that could be obtained in that way, the icosahedral one produces viruses with individual proteins that are to the best approximation equivalent to each other. This is known as the principle of quasiequivalence. Intriguingly, similar mathematics also governs the structural principles behind the construction of geodesic domes popularized by Buckminster Fuller [2,3] and the so-called giant icosahedral fullerenes [4]. It is known today that not all viruses can be fitted in the simplest Caspar-Klug quasiequivalence scheme [5], but some form of mathematical regularity in the organization of proteins of non-Caspar-Klug viruses is often evident [6–8].

The assembly of viruses is more and more being thought of as a process not too different from the cluster nucleation [9] that was known for a long time in several different contexts [10,11]. It was more than fifty years ago since Fraenkel-Conrat and Williams demonstrated that fully infectious tobacco mosaic viruses could be created simply by mixing the viral RNA molecules together with the viral proteins [12]. Under the right conditions (pH and salinity), the viruses

formed spontaneously, i.e., without any special external impetus. This suggests that the process of spontaneous assembly of simple viruses can be understood by relatively simple thermodynamics, presuming that energetics and entropy of assembly can be provided by some convenient model. Indeed, that line of reasoning is often applied in the studies of viruses that proceed from a physical point of view [13]. Not all viruses self-assemble in *in vitro* conditions, but many simple viruses containing ssRNA molecule do. These viruses are the focus of our study.

The physical approach to the self-assembly problem consists, of course, in pinpointing the features of the assembly that are generic and nonspecific and that can be applied with certainty to a large class of viruses. It has been demonstrated that electrostatic interactions play an important role in that respect [14–17]. In particular, the assembly of empty viral capsids that does take place in certain range of solution salinity and pH has been explained in terms of competing attractive hydrophobic and repulsive electrostatic interactions acting between the proteins of a virus capsid [14,15]. It seems natural to extend these studies so that they include the attractive electrostatic interactions acting between the highly negatively charged viral ssRNA molecule (one elementary charge per nucleotide) and the dominantly positively charged capsid. And indeed, this has been done [16–19], but two detailed studies performed thus far [16,17] have reached different conclusions concerning the thermodynamically optimal composition of a virus. While Belyi and Muthukumar predicted that the total charge on the capsid and on the ssRNA in the optimal case are to an excellent approximation equal [16], van der Schoot and Bruinsma found out that, in the optimal case, the total charge on the ssRNA is about twice the charge on the capsid. It is of interest to re-examine these studies, especially in light of recent experimental detection of synthetic viral-like particles composed of viral proteins of cowpea chlorotic mottle virus that encapsidate a poly(styrene sulfonate) molecule, whose total charge is nine

\*asiber@ifs.hr

times larger than the charge on the capsid [20]. There is an obvious need for removing discrepancies between different models of electrostatically guided spontaneous assembly of viruses, which is a basic motivation for our study.

In Sec. II, we present a theoretical model for the free energy of the assembled capsids. The model includes the entropic contributions of salt ions and the encapsidated polyelectrolyte, the nonelectrostatic hardcore self-interaction of the polyelectrolyte and the electrostatic self-interactions of the polyelectrolyte as well as its electrostatic interactions with the oppositely charged capsid wall. The model is solved on the mean-field level leading to an appropriate generalized polyelectrolyte Poisson-Boltzmann description. We show how the free energy functional can be used to study thermodynamical preference towards formation of empty or filled capsids. In Sec. III, the model is numerically solved and the results for capsid-ssRNA complexation energetics, depending on the ssRNA length and salt concentration, are presented. We also calculate the spatial density profile of the encapsidated ssRNA and the salt ions. The results are obtained for two different models of capsid charge distribution: In the first one the capsid is represented as an infinitely thin shell of positive charge, and in the second one the charge distribution is assumed spatially extended, accounting for the charged protein tails (or arms) that protrude into the capsid interior. These tails have been identified as an important structural feature in many types of viruses [16]. The pressure that acts on the capsid wall due to its interaction with the ssRNA is also calculated. We find that it acts as to decrease the capsid radius, at least for viruses that are predicted to

assemble spontaneously by our model. Section IV discusses the pros and cons of our study in its application to real viruses. Both the advantages and the limitations of our model are clearly stated.

## II. DEVELOPMENT OF THE MODEL: COMPLEXATION OF EMPTY CAPSIDS AND THE ssRNA MOLECULE

We approach the problem by representing the viral ssRNA as a generic flexible polyelectrolyte with effective monomer size  $a$ , and  $pe$  charge per monomer, where  $e$  is the electron charge and  $0 < p < 1$ . The polyelectrolyte concentration, written as  $\Psi(r)^2$ , and the electrostatic potential  $\Phi(r)$  are treated as continuous real-valued fields that minimize the mean-field free energy functional of the polyelectrolyte-capsid-salt system  $F$  in the subspace of fixed total number of polyelectrolyte monomers  $N$  so that

$$F(N) = \int f(r) d^3r - \mu \left[ \int d^3r \Psi(\mathbf{r})^2 - N \right], \quad (1)$$

where  $\mu$  is the Lagrange multiplier enforcing the condition of fixed number of monomers. The free energy density  $f(r)$  is given within the mean-field ansatz as a sum of polyelectrolyte entropy term, polyelectrolyte nonelectrostatic self-interaction term, electrostatic energies of polyelectrolyte segments, fixed charges residing on the capsid, and salt ions, as well as finally the entropies of the salt ions. It has the standard form [21,22]

$$f(r) = k_B T \left\{ \frac{a^2}{6} [\nabla \Psi(r)]^2 + \frac{v}{2} \Psi(r)^4 \right\} + [ec^+(r) - ec^-(r) - pe\Psi(r)^2 + \rho_p(r)]\Phi(r) - \frac{\epsilon_0 \epsilon}{2} [\nabla \Phi(r)]^2 + \sum_{i=\pm} \{k_B T [c^i(r) \ln c^i(r) - c^i(r) - (c_0^i \ln c_0^i - c_0^i)] - \mu^i [c^i(r) - c_0^i]\}. \quad (2)$$

Here  $T$  is the temperature,  $k_B$  is the Boltzmann constant,  $c^\pm$  are the concentrations of  $+$  and  $-$  monovalent salt ions, with  $c_0^\pm$  being their bulk concentrations, and  $\mu^\pm$  their chemical potentials,  $\epsilon\epsilon_0$  is the permittivity of water, and  $v$  is the (nonelectrostatic) excluded volume of the polyelectrolyte chain. The mean-field free energy was written within the ground state dominance ansatz, often used for neutral polymers and valid for long polymer chains [23,24]. The solvent contribution to the free energy is effectively described only in the excluded volume term. Explicit solvent contributions, such as discussed recently by Kumar and Muthukumar [25], are assumed to be subdominant to the electrostatic contribution arising from the interaction of oppositely charged capsid and ssRNA chain. The density of charge located on the capsid proteins is denoted by  $\rho_p(r)$ , so that the total capsid charge is  $Q_c = \int d^3r \rho_p(r)$ . We shall consider this charge density to be fixed, i.e., we shall investigate the issues of stability and energetics with respect to the reference state that consists of

the already *assembled* capsids at fixed (physiological) pH.

The variation of the free energy functional with respect to fields  $\Psi$ ,  $\Phi$  and  $c^\pm$  yields two coupled nonlinear equations—the generalized polyelectrolyte Poisson-Boltzmann equation [26],

$$\epsilon_0 \epsilon \nabla^2 \Phi(r) = 2ec_0 \sinh[\beta e \Phi(r)] - ep\Psi(r)^2 + \rho_p(r), \quad (3)$$

and the Edwards equation [27],

$$\frac{a^2}{6} \nabla^2 \Psi(r) = v\Psi(r)^3 + pe\beta\Phi(r)\Psi(r) - \mu\Psi(r). \quad (4)$$

These two coupled nonlinear equations should be solved requiring that the polyelectrolyte density amplitude field vanishes at the interior capsid radius,  $R$ . These equations represent a generalization of the self-consistent field equations of neutral polymers [23].

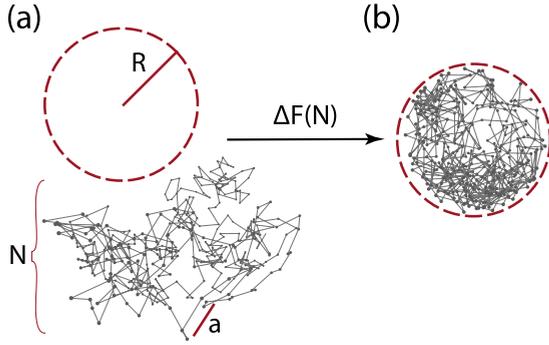


FIG. 1. (Color online) Panel (a) represents the reference state, i.e., the empty capsid and the dissolved ssRNA, while panel (b) represents the filled capsid that contains the ssRNA. The free energies of the reference and the filled states are  $F(0)$  and  $F(N)$ , respectively [see Eq. (5) and the discussion following it].

It should be reiterated that the free energy functional in Eq. (2) corresponds to the already assembled spherical capsids of radii  $R$  with protein charge given by  $\rho_p(r)$ , and containing  $N$  ssRNA bases. As such, this functional can not yield answers regarding the thermodynamics of complexation of disassembled capsids (individual proteins) and dissolved ssRNA molecule. However, a simple functional constructed as

$$\Delta F(N) = F(N) - F(0), \quad (5)$$

can be used to discuss the thermodynamic preference towards the formation of filled or empty protein capsids, since  $F(0)$  is simply the free energy functional of the empty capsid. Presuming that the free energy of ssRNA in the bulk solution is vanishingly small,  $\Delta F(N)$  can be considered as the free energy functional describing the complexation of the capsids with the ssRNA molecule containing  $N$  monomers, as indicated schematically in Fig. 1. A vanishingly small contribution of the dissolved ssRNA is in fact a direct consequence of the ground state dominance ansatz that was invoked in the construction of Eq. (2) and can thus be thought of as a feature of our theoretical framework (see, however, Ref. [28]). On the level of the ground state dominance, the polyelectrolyte free energy is zero just by sheer normalization of the polymer density field. This corroborates the identification of  $\Delta F$  with the capsid-ssRNA complexation functional. The physical content of  $F(0)$ , i.e., the empty capsid free energy, is an electrostatic repulsion of the proteins that comes out directly from Eq. (2) plus the attractive protein-protein interactions that are not defined by Eq. (2) but which influence only the assembly of individual proteins in a capsid and do not contribute to the empty capsid-ssRNA complexation energetics [the attractive part of the protein-protein interactions that is missing from  $F(N)$  and  $F(0)$  as defined by Eq. (2) cancels out in  $\Delta F$ ]. Note that  $F(0)$  depends on the salt concentration due to its electrostatic component—this was investigated in detail in Ref. [15].

### III. NUMERICAL RESULTS FOR CAPSID-SSRNA COMPLEXATION

Since the size of empty capsids is usually the same as in the fully functional viruses (at least in a range of pH and

salinity values [29,30]), it is reasonable to fix the capsid radius at a prescribed value, corresponding to the preferred mean curvature of the empty capsid [15]. There are three remaining parameters that importantly influence the energetics of the capsid-ssRNA complexation. These are the ssRNA length (i.e., the number of monomers  $N$ ), the bulk concentration of monovalent salt  $c_0$  and the density of capsid charge  $\rho_p(r)$ , which depends on the aminoacid composition and shape of an individual protein in a capsid. The numerical method that we use for solving equations (3) and (4) consists of representing the fields  $\Psi(r)$  and  $\Phi(r)$  on a mesh of radial points and treating the two equations as a set of coupled nonlinear algebraic equations for finite number of values of  $\Psi(r)$  and  $\Phi(r)$  at the mesh nodes. The thus obtained set of equations is solved using Levenberg-Marquardt algorithm [31]. We employed a similar numerical method in Ref. [15].

#### A. Thin capsid

We shall first examine the simplest possible case, i.e. when the capsid is assumed to be infinitely thin and the protein charge is distributed uniformly over the sphere of radius  $R$ , so that the surface charge density is  $\sigma$ . This is the case studied in Ref. [17]. We shall first calculate the free energies of the polyelectrolyte-capsid complexes. It is of interest to separate the entropic ( $F_{\text{ent}}$ ) and nonelectrostatic (hardcore) self-interaction ( $F_{\text{int}}$ ) contributions of the polyelectrolyte

$$F_{\text{ent}}(N) = k_B T \frac{a^2}{6} \int d^3 r [\nabla \Psi(r)]^2, \quad (6)$$

$$F_{\text{int}}(N) = k_B T \frac{v}{2} \int d^3 r \Psi(r)^4,$$

so to study the interplay of entropic and enthalpic contributions to the total free energy. Figure 2 displays the total free energy  $F(N)$  and  $F_{\text{ent}}(N)$  and  $F_{\text{int}}(N)$  contributions for three different salt concentrations. One can note that the nonelectrostatic self-repulsion of the polyelectrolyte is almost negligible for our choice of the excluded volume parameter. The polyelectrolyte entropic contribution is, however, quite important reaching  $\sim 25\%$  for  $c_0 = 10$  and 100 mM salt concentrations.

Figure 3 displays complexation energies as a function of  $N$  for three different salt concentrations as denoted in the figure. Several important messages can be read off directly from Fig. 3. First, there is a critical number of monomers ( $N_c$ ) that can be thermodynamically packed within a capsid. This happens at the point where  $\Delta F(N_c) = 0$ , since for all  $N > N_c$ ,  $\Delta F$  is positive which means that the formation of a phase containing empty capsids and freely floating ssRNA is preferred thermodynamically over a phase containing filled viruses. The critical number of monomers is in this case strongly influenced by the salt concentration, and for sufficiently large salinity ( $c_0 \sim 800$  mM for the choice of parameters in Fig. 3),  $N_c$  approaches zero which means that even the shortest ssRNA molecule cannot be packed within the virus. Note, however, that in that case our continuum model

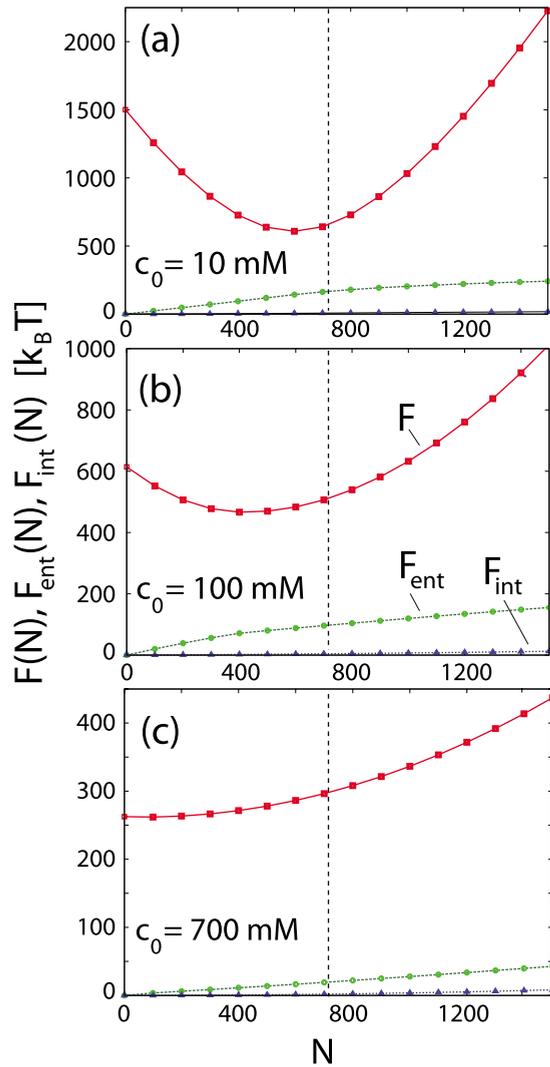


FIG. 2. (Color online) Total [ $F(N)$ , squares], entropic [ $F_{\text{ent}}(N)$ , circles], and polyelectrolyte self-interaction [ $F_{\text{int}}(N)$ , triangles] free energies of the polyelectrolyte-capsid-salt system as functions of the number of monomers. The parameters used are  $p=1$ ,  $a=0.5$  nm,  $v=0.05$  nm<sup>3</sup>,  $T=300$  K,  $\sigma=0.4$  e/nm<sup>2</sup>, and  $R=12$  nm. The results are shown for three different bulk concentrations of salt ions,  $c_0=10$  [panel (a)], 100 [panel (b)], and 700 mM [panel (c)]. The vertical dashed line denotes the number of monomers for which the polyelectrolyte charge equals in magnitude the capsid charge ( $N \approx 724$ ).

cannot be applied with certainty. Nevertheless, the results are highly suggestive and they predict that increased salt concentration hinders the formation of filled capsids, i.e., functional viruses. The maximal length of ssRNA that can be thermodynamically packed in the capsid without producing an empty capsid is such that the charge on the ssRNA is twice the charge on the capsid, but this can take place only in extremely low salt solutions ( $c_0 \sim 1$  mM) where the capsid alone may be destabilized [15]. Second, the thermodynamically optimal number of monomers ( $N_0$ ) for which  $\partial\Delta F/\partial N=0$  is also strongly influenced by salt, again becoming smaller as the salt concentration increases. The optimal number of monomers in low salt solutions is such that the

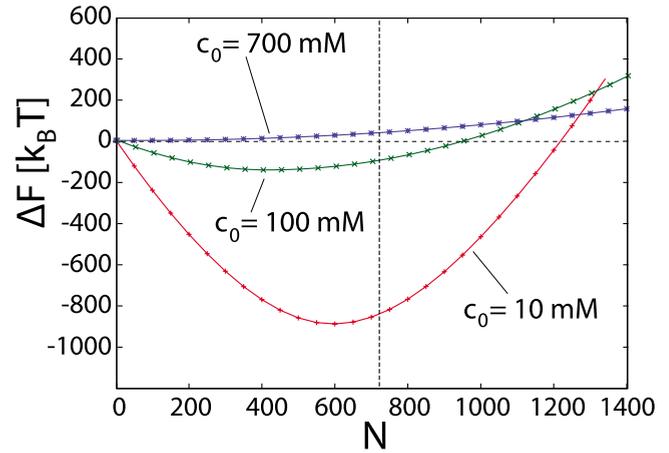


FIG. 3. (Color online) The complexation free energies of the polyelectrolyte-capsid-salt system as functions of the number of monomers. The parameters used are  $p=1$ ,  $a=0.5$  nm,  $v=0.05$  nm<sup>3</sup>,  $T=300$  K,  $\sigma=0.4$  e/nm<sup>2</sup>, and  $R=12$  nm. The results are shown for three different bulk concentrations of salt ions,  $c_0=10$ , 100, and 700 mM. The vertical dashed line denotes the number of monomers for which the polyelectrolyte charge equals in magnitude the capsid charge ( $N \approx 724$ ).

total (negative) charge on the ssRNA equals the (positive) charge on the capsid. In contrast to Ref. [17], we find that the optimal number of monomers is such that the magnitude of charge on the polyelectrolyte is always smaller than the protein charge (summing only the polyelectrolyte and capsid charges, we find that the total virus charges at the optimal polyelectrolyte lengths are +120, +275, and +675  $e$  for  $c_0=10$ , 100, and 700 mM, respectively). Note, however, that the whole system, i.e., the polyelectrolyte, the capsid, and the salt ions is always electroneutral. Third, the complexation is more efficient in solutions with smaller salt concentrations which is indicated by deepening of the minimum in  $\Delta F$  [ $\Delta F(N_0)$ ] as the salt concentration decreases. However, for small salt concentrations the formation of capsids may be inhibited due to the increased electrostatic repulsion between the capsid proteins [15]. In fact, looking only at  $F(N)$  as defined in Eq. (2), one finds that it diminishes with the increase of salt concentration for all  $N$  (see Fig. 2).

Our approach is constructed to not only yield free energies of complexation, but also the polyelectrolyte density and electrostatic potential fields which are in fact the primary unknowns. The polyelectrolyte density is a quantity that can be measured and it is of interest to investigate it in more details. In Fig. 4, we present how the density changes with the number of monomers for fixed salt concentration  $c_0=100$  mM.

Note that the ssRNA is always concentrated in a shell close to the capsid. This is in accord with several experimental determinations of the RNA density, including the cases of RNA containing hepatitis B precursor viruses [32], flock house virus [33] that contains double stranded RNA, and rice yellow mottle and southern bean mosaic RNA viruses [34]. We have found that the effective thickness of the high-concentration ssRNA shell changes both with the salt concentrations (becoming smaller in low salt, consistent with the

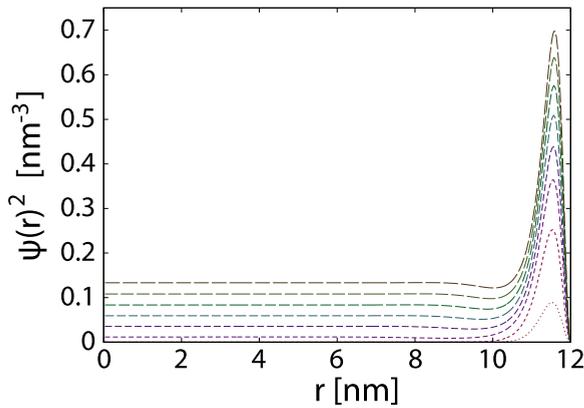


FIG. 4. (Color online) The polyelectrolyte concentration profile for  $c_0=100$  mM. The curves displayed correspond to  $N=100, 300, 500, 700, 900, 1100, 1300,$  and  $1500$ . The lines are styled so that the length of their dashes is proportional to  $N$ . Other parameters of the calculation are the same as in Fig. 3.

decrease of Debye-Hückel screening length [15]) and the excluded volume parameter (becoming larger as the excluded volume increases). An interesting feature of these calculations is that they predict a partial filling of the capsid interior (core region of small radial values) for superoptimal monomer numbers (in this case  $N_0 \sim 450$ ). This effect becomes progressively more pronounced as  $N$  increases to a larger extent over  $N_0$ . This is somewhat similar to the case of sindbis virus, where it was experimentally observed that there is a maximum in the RNA density in the region where the RNA contacts the capsid proteins, in addition to the uniformly filled viral core [35]. Our results indicate that this may be a virus with superoptimal RNA length. The radial distribution of densities of counterions can be obtained easily from the electrostatic potential. This is shown in Fig. 5 for the same conditions as in Fig. 4.

The radial profiles of counterion densities for  $N=700$  (close to charge neutrality point) are somewhat similar to those obtained in Ref. [36], where the authors used Monte Carlo techniques to solve the discrete version of a model similar to ours (one should note, however, that the diameter of the capsid studied there is more than twice smaller than ours). Note that for superoptimal polyelectrolyte lengths, the density of positive counterions in the capsid core (region of small radii) is larger from the bulk density, which is a reflection of the fact that the superoptimal polyelectrolyte fills the capsid core and the counterions accumulate there to screen it.

### B. Capsid with protein tails protruding into its interior

Not all ssRNA viruses can be represented by an infinitely thin shell of positive protein charge density. While this may be a reasonable approximation for viruses as dengue or yellow fever, it is certainly a poor approximation for, e.g., cucumber mosaic virus (see Fig. 6), tomato aspermy virus and the much investigated cowpea chlorotic mottle virus [37]. These viruses are known to have specifically shaped capsid proteins, so that their  $N$ -terminal tails are highly positively charged and stretched. When a virus of this type is fully

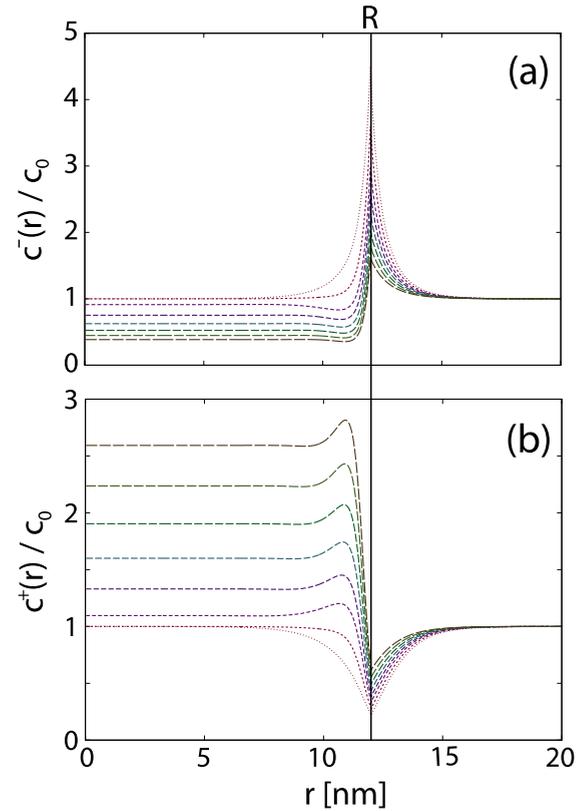


FIG. 5. (Color online) (a) Density of negative monovalent counterions. (b) Density of positive monovalent counterions. The bulk salt concentration is  $c_0=100$  mM. The curves displayed correspond to different polyelectrolyte lengths ( $N=100, 300, 500, 700, 900, 1100, 1300,$  and  $1500$ ). The lines are styled so that the length of their dashes is proportional to  $N$ . Other parameters of the calculation are the same as in Fig. 3. Thick vertical line denotes the position of the capsid (capsid radius  $r=R$ ).

assembled, the capsid protein  $N$ -tails protrude into its interior as can also be seen in Fig. 6. It has recently been suggested that the existence of highly basic capsid peptide tails can explain the proportionality between the net charge on the capsid proteins and the total length of the ssRNA viral genome [16]. Here we are more interested in investigating whether such delocalization of the protein charge influences the complexation of a virus. We represent the capsid charge density as

$$\rho_p(r) = \frac{Q_c}{4\pi r^2 \xi}, \quad R - \xi < r < R, \quad (7)$$

and  $\rho_p(r)=0$  otherwise, i.e., we treat the capsid peptide arms as strongly stretched polyelectrolytes, but not necessarily of brush type [16], of length  $\xi$ , carrying in total a charge  $Q_c$  per capsid (for  $\xi \ll R$ ). The illustration of the geometrical parameters of our model density is shown in Fig. 6 on the experimentally determined structure of a cucumber mosaic virus capsid.

The actual charge distribution in real viruses depends on the amino acid content of the capsid peptide tails, but it is worth mentioning that results practically indistinguishable

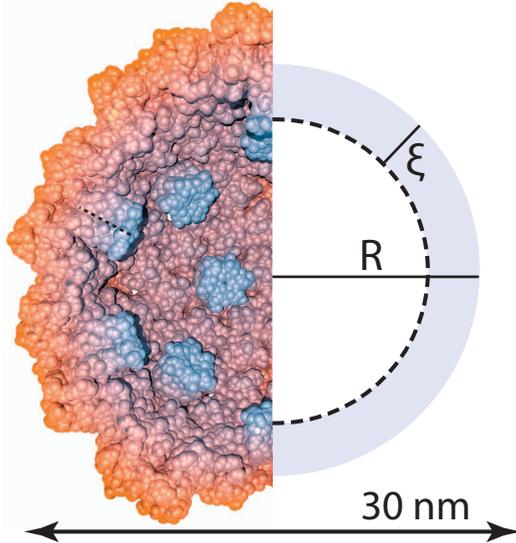


FIG. 6. (Color) One quarter of the cucumber mosaic virus capsid (strain FNY). The image was constructed by applying the group of icosahedral transformations to the RCSB Protein Databank [38] entry 1F15 and all atoms in the resulting structure were represented as spheres of radius  $3.4 \text{ \AA}$  (which is the experimental resolution [37]). They were colored in accordance with their distance from the geometrical center of the capsid, so that the atoms that are farthest away from the center are orange, while those that are closest to the center (and belonging to the capsid protein tails) are light blue.

from those shown in Fig. 7 are obtained by assuming that  $\rho_p = \text{const}$  for  $R - \xi < r < R$  and zero otherwise, so that the total protein charge is still  $Q_c$  (the robustness of these results is mostly due to the fact that  $\xi \ll R$ ). Note that we do not account for steric repulsion acting between the capsid peptide arms and the viral ssRNA, i.e., for the loss of interior capsid volume available to ssRNA resulting from the protrusion of pieces of capsid proteins into the capsid interior. Figure 7 displays the complexation free energies for the capsid density modeled by Eq. (7).

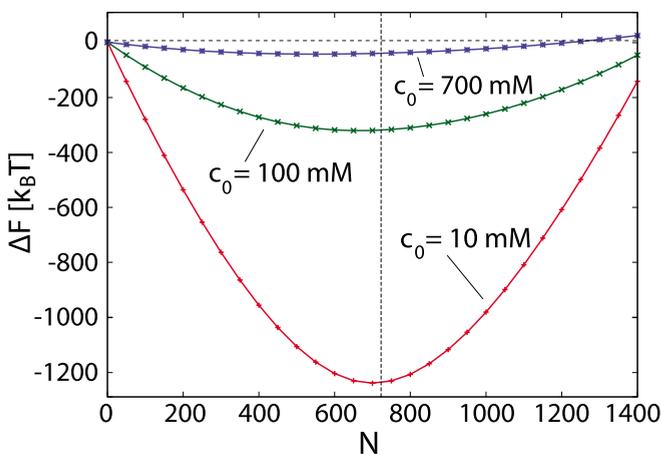


FIG. 7. (Color online) The complexation free energies for the capsid charge modeled as in Eq. (7) with  $Q_c = 724 e$  and  $\xi = 2.5 \text{ nm}$ . The other parameters of the calculation are the same as in Fig. 3.

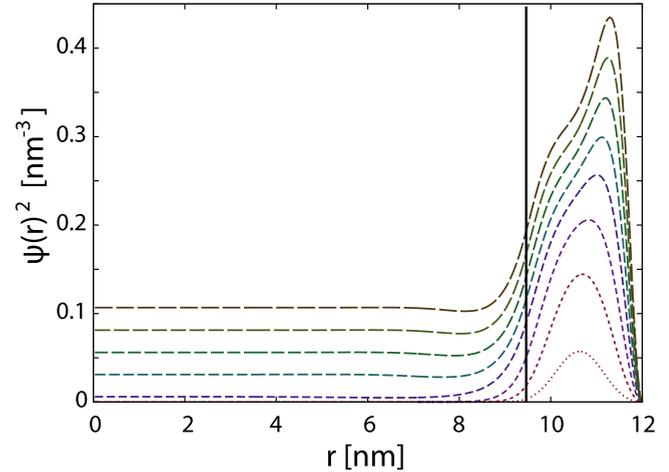


FIG. 8. (Color online) The polyelectrolyte concentration profile for  $c_0 = 100 \text{ mM}$  and for the capsid density model represented in Eq. (7) and used in the calculation in Fig. 7. The curves displayed correspond to  $N = 100, 300, 500, 700, 900, 1100, 1300,$  and  $1500$ . The lines are styled so that the length of their dashes is proportional to  $N$ . A vertical line at  $r = 9.5 \text{ nm}$  indicates the position of the  $N$ -termini of the capsid proteins. The other parameters of the calculation ( $p, a, v, T,$  and  $R$ ) are the same as in Fig. 3.

There are several important differences with respect to the infinitely thin capsid model. The complexation is more robust with respect to the changes in salt concentration. This is seen by a much lesser influence of the salt concentration on the optimal monomer number (for  $c_0 = 10, 100,$  and  $700 \text{ mM}$ , the optimal monomer numbers are  $700, 650,$  and  $550$ , respectively, so that the total virus charge is  $+25, +75,$  and  $+175 e$ ). This is in rough agreement with the findings by Belyi and Muthukumar who estimate that the total ssRNA charge and the total capsid charge are equal up to a quantity of the order of 10 elementary charges for salt concentrations below about  $c_0 = 100 \text{ mM}$  [16]. Their model also includes the protein charge that is delocalized on the protein tails protruding into the capsid interior. Our results thus bridge the two [16,17] apparently contradictory previous attempts to describe the viral energetics and show that the spatial distribution of protein charge determines the important features of the energetics of viruses with regard to salt concentration. We conclude that the delocalization of the charge density on the protein tails may contribute to the robustness of the viral assembly and we speculate that it may offer an evolutionary advantage to such viruses. Although we see the effect of the tails only in the complexation energetics, it is reasonable to assume that they will also speed up the kinetics of assembly (see also Ref. [39] for the study of this effect). This can be inferred from our data from the deeper minima of  $\Delta F$  with respect to the case of the infinitely thin capsid—note that the total capsid charge is the same in both cases (for  $c_0 = 100 \text{ mM}$ , the difference in complexation free energies is about  $150 k_B T$  per capsid at the optimal monomer number). Due to the polyelectrolyte delocalization (see Fig. 8), its entropic contribution to the free energy is significantly smaller in the capsid with tails. In particular, we have found that the entropic contribution of the polyelectrolyte with optimal length (for  $c_0 = 100 \text{ mM}$ ) is about four times smaller than in

the case of the infinitely thin capsid ( $\sim 20k_B T$  vs  $\sim 80k_B T$ —this accounts for about 40% of the difference in the complexation free energies). We have numerically found that the complexation energetics depends on the tail lengths  $\xi$ , but that even for quite short tails,  $\xi=1.5$  nm, the complexation energetics is improved and yields results that are closer to those presented in Fig. 7 than in Fig. 3. Our findings can be easily understood in the context of the Debye-Hückel approximation, since the delocalization of the capsid charge on the tails makes the screening of attractive ssRNA-protein electrostatic interaction by salt less efficient. As already mentioned, the capsid protein tails also influence the density of the encapsidated ssRNA. This is shown in Fig. 8.

Note that the thickness of the ssRNA “shell” is now determined by the length of the protein arms, unlike in the case of infinitely thin shell of viral protein charge where it is determined by  $a$  and  $v$  parameters of the polyelectrolyte. For superoptimal polyelectrolyte lengths there appears a characteristic “two-humped” profile of the polyelectrolyte concentration which results from an interplay of the four length scales involved in this case—the  $a$  and  $v$  parameters of the polyelectrolyte, the length of the capsid arms,  $\xi$  and the ratio of the capsid radius and the Debye-Hückel screening length that depends on the salt concentration.

### C. Osmotic pressure in ssRNA viruses

The osmotic pressure in ssRNA viruses has been recently much debated [17,40] since it could have some consequences on the release of viral genome in the cell. It is known that the double-stranded DNA viruses are under a large positive osmotic pressure ( $\sim 50$  atm) [41,42] acting from the inside out, so to pack their genome they need special strategies (molecular motors that use up ATP molecules). On the other hand, ssRNA viruses self-assemble without any biochemical fuel consumption and it has been suggested that this may indicate that the osmotic pressure within them is close to zero [40]. The calculation of osmotic pressure is complicated by an attractive contribution (of nonelectrostatic origin) to the virus energetics, not accounted for by the results shown in Fig. 3. This is the contribution that keeps the empty capsids together when they form in the absence of the polyelectrolyte [15]. We shall assume that the derivative of this contribution with respect to the capsid radius exactly cancels the analogous derivative of electrostatic self-repulsion of the empty capsid. The reasonableness of this assumption is corroborated by the fact that the empty capsids form preferably at a single radius which is in most cases the radius of the fully functional virus [29,30]. Alternatively, our results can be interpreted as the excess osmotic pressures induced by electrostatic interactions  $\Pi$  with respect to those acting in the empty capsids, viz.

$$\Pi(N) = -\frac{1}{4\pi R^2} \left( \left. \frac{\partial F}{\partial R} \right|_{Q_c, N} - \left. \frac{\partial F}{\partial R} \right|_{Q_c, N=0} \right). \quad (8)$$

The osmotic pressures defined in this way are calculated for the model of infinitely thin charged capsid and shown in Fig. 9.

Intriguingly, in the range of polyelectrolyte lengths for which the filled viruses are more stable than the empty ones,

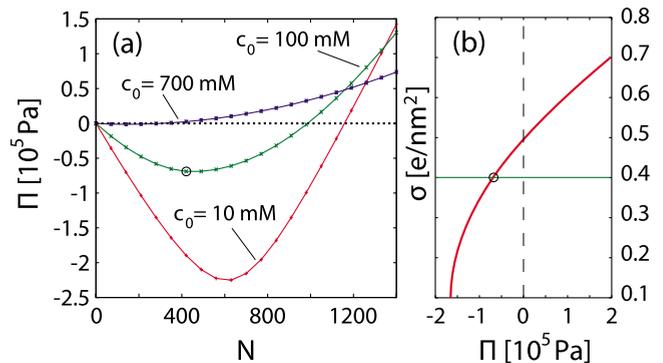


FIG. 9. (Color online) (a) Osmotic pressure acting on the viral capsid as a function of the number of monomers. (b) Osmotic pressure for fixed number of monomers  $N=420$ ,  $c_0=100$  mM [the point denoted by a circle in panel (a)] as a function of the capsid charge density  $\sigma$ . The other relevant parameters of the calculation are the same as in Fig. 3.

osmotic pressures are negative (inward), i.e., the electrostatic forces act to decrease the radius of the capsid. Osmotic pressures vanish close to the border of feasibility of spontaneous self-assembly of filled capsids and change the sign afterwards. The typical magnitudes of the pressures are about 0.5 atm at physiological salt conditions, but we have found even smaller pressures for capsids of larger radii. Very similar results are also found for the capsid with the charges delocalized on the protein tails. These results are in clear contrast with those obtained previously for double-stranded bacteriophages, where outward pressures two orders of magnitude larger than what we find here for ssRNA viruses, have been calculated [41–43].

The negative values of osmotic pressure are commonly found in the interaction of charged planar surfaces with an oppositely charged polyelectrolyte chain in between [44]. In that case they are due to bridging configurations of the chain, inducing attractive interactions between apposed surfaces that result in negative osmotic pressures [26]. One is thus tempted to connect the negative values of the osmotic pressure, in the case discussed here, with a similar mechanism. The ssRNA enclosed within the capsid would thus bridge the space between neighboring sections along the curved capsid and induce attractive interactions between them. Summing these attractions along the total surface of the capsid would give rise to an overall negative osmotic pressure of the polyelectrolyte. This interpretation is certainly corroborated by the characteristic “humped” profile of the polyelectrolyte concentration in the vicinity of the capsid wall, which is in fact very similar to the bimodal polyelectrolyte concentration profile observed in the case of planar confinement [26].

It is known that changes of pH may induce the disassembly of viruses. Assuming that such changes modify the capsid charge density, it is of interest to investigate how the osmotic pressure acting on an assembled capsid changes as  $\sigma$  is varied. This information is represented in Fig. 9(b) for  $c_0=100$  mM and  $N=420$ . The protonation of the capsid proteins [45] may indeed result in positive osmotic pressures which could play a role in capsid disassembly in both *in vitro* and *in vivo* conditions. An increase of about 40% in total

protein charge is needed to revert the sign of the pressure. On the other hand, a decrease of salt concentration would induce an enhanced inward pressures on viral capsids, which may be an alternative pathway for viral disassembly.

#### IV. PROS AND CONS FOR APPLICATION OF THE PROPOSED MODEL TO REAL VIRUSES

The approach that we developed in this paper is in several respects a refined version of previous theoretical approaches. In particular, once we adopt the model of the mean-field free energy in Eq. (2), we can (numerically) exactly follow its thermodynamical consequences. However, the way that the model was set up dictates the choice of the reference state that may not be the one that yields the most insight. One would certainly prefer the reference state that consists of the dissolved proteins and the ssRNA. To model the self-assembly from this starting point would, however, require (at least) a model of the protein-protein attractive interactions which are not very well known and not readily available [14,45]. Nevertheless, our study predicts the preference towards formation of empty capsids or functional viruses and details how it depends on the electrostatic interactions and entropic contributions related to encapsidated ssRNA. This appears to us to be a very useful result.

The proposed model free energy is approximate in several respects. First, the model is a continuum one, and as such it is strictly appropriate for the study of sufficiently long ssRNA. Second, spherical symmetry of the polyelectrolyte density and the ensuing electrostatic potential is assumed *a priori*. It is well known that the asphericity of “spherical” (icosahedral) viruses increases with the mean radius of the virus [46]. However, deviations from the perfectly spherical shape are generally small even for quite large viruses [46,47], thus our approximation of spherical symmetry is not expected to be a serious limitation. Third, the model does not account for mechanical elasticity of the polyelectrolyte, i.e., our approach is tailored to flexible polyelectrolyte molecules, the ssRNA in particular. Our continuum approach also can not account for the details of the RNA conformation, such as its branched and locally double stranded structure, or its possible dodecahedral ordering in the vicinity the capsid, a feature that has been theoretically investigated recently [40,48] and also found experimentally [49]. Fourth, the ground state dominance approximation for the polyelectrolyte density profile has been adopted, which additionally limits the application of our study [50]. Again, the study is best suited for sufficiently long polyelectrolytes. Fifth, the salt is treated on

the Poisson-Boltzmann mean-field level, which means that the model is best suited for monovalent salt. And finally, solvent effects are treated only effectively via a nonelectrostatic polymer self-interaction term. In the case of polyelectrolyte confinement in noncharged enclosures the explicit solvent effects have recently been found to be important [25]. It nevertheless seems reasonable that they are subdominant to electrostatics in the case of confinement within a charged boundary with a charge opposite to the one residing on the polyelectrolyte chain.

Above all of these approximations is a neglect of any chemical specificity in ssRNA-protein interactions. These could be approximately added to the free energy functional as a fixed capsid binding contribution depending simply on the number of capsid proteins. This term would, of course, reflect in the free energy  $F(N)$  of assembled capsids and not in the reference state  $F(0)$ . Thus, the complexation energy  $\Delta F(N)$  would be simply shifted towards more negative values. This effect could in fact explain the spontaneous assembly of viruses whose ssRNA charge is more than twice larger than the capsid charge. According to our calculation, this is the only way to explain a significant negative overcharging that may be present in some viruses [40], that was not, however, found for the cases studied in Ref. [16]. Our study does predict that the spontaneous assembly of negatively charged (but nonoptimal) viruses is possible, but only if its total (negative) charge is smaller in magnitude from the (positive) capsid charge. For viruses with charged protein tails, this scenario is possible even in physiological salt concentrations, see Fig. 7. However, it could be speculated that small ssRNA viruses that rely exclusively on the spontaneous assembly could in fact have too small positive charge to pack the whole genome that typically contains the information on the capsid proteins and additional proteins used to replicate the viral RNA. In this context, it is interesting to note that there are ssRNA viruses (bromoviruses,  $R \sim 14$  nm, see, e.g., Ref. [51]) that do not pack the whole genome within the capsid, but rather pack pieces of the RNA in separate viral particles. Our results suggest that this could be a consequence of small effective charge on the viral proteins, or too large charge contained in the (complete) viral RNA sequence.

#### ACKNOWLEDGMENTS

A.Š. acknowledges support by the Ministry of Science, Education, and Sports of Republic of Croatia through Project No. 035-0352828-2837. R.P. acknowledges support by the Agency for Research and Development of Slovenia under Grant No. P1-0055(C).

[1] D. L. D. Caspar and A. Klug, Cold Spring Harbor Symp. Quant. Biol. **27**, 1 (1962).  
 [2] A. Šiber, e-print arXiv:0711.3527.  
 [3] G. J. Morgan, Trends Biochem. Sci. **28**, 86 (2003).  
 [4] A. Šiber, Nanotechnology **17**, 3598 (2006).  
 [5] T. S. Baker, N. H. Olson, and S. D. Fuller, Microbiol. Mol.

Biol. Rev. **63**, 862 (1999).  
 [6] E. Kellenberger, Eur. J. Biochem. **190**, 233 (1990).  
 [7] V. L. Lorman and S. B. Rochal, Phys. Rev. Lett. **98**, 185502 (2007).  
 [8] R. Twarock, J. Theor. Biol. **226**, 477 (2004); R. Twarock, Philos. Trans. R. Soc. London, Ser. A **364**, 3357 (2006).

- [9] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, *Biophys. J.* **90**, 1939 (2006).
- [10] J. Groenewold and W. Kegel, *J. Phys. Chem. B* **105**, 11702 (2001).
- [11] J. Groenewold and W. Kegel, *J. Phys.: Condens. Matter* **16**, S4877 (2004).
- [12] H. Fraenkel-Conrat and R. C. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **41**, 690 (1955).
- [13] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, *Phys. Rev. Lett.* **90**, 248101 (2003).
- [14] W. K. Kegel and P. van der Schoot, *Biophys. J.* **86**, 3905 (2004).
- [15] A. Šiber and R. Podgornik, *Phys. Rev. E* **76**, 061906 (2007).
- [16] V. A. Belyi and M. Muthukumar, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17174 (2006).
- [17] P. van der Schoot and R. Bruinsma, *Phys. Rev. E* **71**, 061928 (2005).
- [18] T. Hu, R. Zhang, and B. I. Shklovskii, *Physica A* **387**, 3059 (2008).
- [19] S. I. Lee and T. T. Nguyen, *Phys. Rev. Lett.* **100**, 198102 (2008).
- [20] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. Gelbart, *Biophys. J.* **94**, 1428 (2008).
- [21] I. Borukhov, D. Andelman, and H. Orland, *J. Phys. Chem. B* **103**, 5042 (1999).
- [22] A. Shafir, D. Andelman, and R. R. Netz, *J. Chem. Phys.* **119**, 2355 (2003).
- [23] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University, Ithaca, 1979).
- [24] G. H. Fredrickson, *The Equilibrium Theory of Inhomogeneous Polymers* (Clarendon Press, Oxford, 2006).
- [25] R. Kumar and M. Muthukumar, *J. Chem. Phys.* **128**, 184902 (2008).
- [26] R. Podgornik, *J. Polym. Sci., Part B: Polym. Phys.* **42**, 3539 (2004).
- [27] S. F. Edwards, *Proc. Phys. Soc. London* **85**, 613 (1965).
- [28] A. V. Dobrynin and M. Rubinstein, *Prog. Polym. Sci.* **30**, 1049 (2005).
- [29] K. Iwasaki, B. L. Trus, P. T. Wingfield, N. Cheng, G. Campusano, V. B. Rao, and A. C. Steven, *Virology* **271**, 321 (2000).
- [30] A. Zlotnick, N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, S. J. Stahl, and P. T. Wingfield, *Biochemistry* **35**, 7412 (1996).
- [31] B. S. Garbow, K. E. Hillstrom, and J. J. More, MINPACK library of FORTRAN routines, Argonne National Laboratory, MINPACK Project, 1996.
- [32] A. Zlotnick, N. Cheng, J. F. Conway, A. C. Steven, and P. T. Wingfield, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9556 (1997).
- [33] M. Tihova, K. A. Dryden, T. L. Le, S. C. Harvey, J. E. Johnson, M. Yeager, and A. Schneeman, *J. Virol.* **78**, 2897 (2004).
- [34] N. Opalka, M. Tihova, C. Brugidou, A. Kumar, R. N. Beachy, C. M. Fauquet, and M. Yeager, *J. Mol. Biol.* **303**, 197 (2000).
- [35] W. Zhang, S. Mukhopadhyay, S. V. Pletnev, T. S. Baker, R. J. Kuhn, and M. G. Rossmann, *J. Virol.* **76**, 11645 (2002).
- [36] D. G. Angelescu, J. Stenhammar, and P. Linse, *J. Phys. Chem. B* **111**, 8477 (2007).
- [37] T. J. Smith, E. Chase, T. Schmidt, and K. L. Perry, *J. Virol.* **74**, 7578 (2000).
- [38] <http://www.rcsb.org>
- [39] T. Hu and B. I. Shklovskii, *Phys. Rev. E* **75**, 051901 (2007).
- [40] R. F. Bruinsma, *Eur. Phys. J. E* **19**, 303 (2006).
- [41] A. Evilevitch, M. Castelnovo, C. M. Knobler, and W. M. Gelbart, *J. Phys. Chem. B* **108**, 6838 (2004).
- [42] P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev, and R. Phillips, *Biophys. J.* **88**, 851 (2005).
- [43] A. Šiber, M. Dragar, V. A. Parsegian, and R. Podgornik, *Eur. Phys. J. E* **26**, 317 (2008).
- [44] R. Podgornik and M. Ličer, *Curr. Opin. Colloid Interface Sci.* **11**, 273 (2006).
- [45] W. K. Kegel and P. van der Schoot, *Biophys. J.* **91**, 1501 (2006).
- [46] J. Lidmar, L. Mirny, and D. R. Nelson, *Phys. Rev. E* **68**, 051910 (2003).
- [47] A. Šiber, *Phys. Rev. E* **73**, 061915 (2006).
- [48] D. G. Angelescu and P. Linse, *Phys. Rev. E* **75**, 051905 (2007).
- [49] L. Tang, K. N. Johnson, L. A. Ball, T. Lin, M. Yeager, and J. E. Johnson, *Nat. Struct. Biol.* **8**, 77 (2001).
- [50] G. J. Fleer, M. A. Cohen Stuart, J. M. H. M. Scheutjens, T. Cosgrove, and B. Vincent, *Polymers at Interface* (Chapman & Hall, London, 1993).
- [51] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, *Virology* **277**, 450 (2000).