Impact of a nonuniform charge distribution on virus assembly

Siyu Li,[1] Gonca Erdemci-Tandogan,[1] Jef Wagner,[1] Paul van der Schoot,[2,3] and Roya Zandi[1]
[1]*Department of Physics and Astronomy, University of California, Riverside, California 92521, USA*
[2]*Group Theory of Polymers and Soft Matter, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
[3]*Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands*
(Received 29 March 2017; published 7 August 2017)

Many spherical viruses encapsulate their genomes in protein shells with icosahedral symmetry. This process is spontaneous and driven by electrostatic interactions between positive domains on the virus coat proteins and the negative genomes. We model the effect of the nonuniform icosahedral charge distribution from the protein shell instead using a mean-field theory. We find that this nonuniform charge distribution strongly affects the optimal genome length and that it can explain the experimentally observed phenomenon of overcharging of virus and viruslike particles.

DOI: 10.1103/PhysRevE.96.022401

I. INTRODUCTION

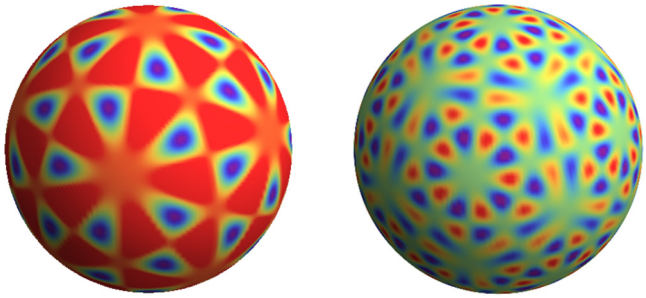
The simplest viruses consist of two components: the genome, either an RNA or DNA polynucleotide that carries the genetic code, and the capsid, a protein shell that encloses the genome. The capsid consists of many identical (or nearly identical) copies of the coat protein subunit. Even though the coat proteins are highly irregular in shape, the protein shells of most spherical viruses are highly structured and obey icosahedral symmetry [1–4]. One of the consequences of icosahedral symmetry is that it puts restrictions on the number of proteins that can make up a spherical virus shell. It limits this number to 60 times the structural index T that almost always assumes certain "magic" integer values $T = 1,3,4,7,\ldots$ [5–7].

Many small single-stranded RNA (ssRNA) viruses have been shown to spontaneously self-assemble *in vitro*, that is, outside living cells in solutions containing virus coat protein subunits and genomes. In fact, virus coat proteins are able to coassemble with a variety of cargos, including RNAs of other and sometimes unrelated viruses, synthetic polyanions, and negatively charged nanoparticles [8–10]. The spontaneous assembly of properly structured viral capsids of many icosahedral RNA viruses with this variety of cargo is due to the presence of a disordered RNA binding domain on the amino-terminal (N-terminal) or the carboxy-terminal (C-terminal) end of the protein subunits. These are rich in basic amino acids that potentially extend quite deep into the capsid interior. These basic amino acids are charged positively under most solution conditions and typically bear a few to tens of positive charges depending on the species of virus. It is now widely accepted that electrostatic interactions between the positive charges on the coat protein tails and negative charges on the genome are the main driving force for the spontaneous assembly of simple viruses in solution [11–19].

Naively, one might expect that the total charge on the genome and the capsid would balance out, if not perfectly, then certainly approximately. However, in many ssRNA viruses the number of negative charges on the genome significantly exceeds the number of positive charges on the capsid proteins (CPs). For example, the number of positive charges on capsids of the Cowpea chlorotic mottle virus (CCMV) and the Brome mosaic virus (BMV), both with $T = 3$ structures, is about 1800, yet their genome measures about 3000 nucleotides (nts) [20]. As each nucleotide bears a single charge, this suggests an *overcharging* of over 60%. Furthermore, in a recent set of *in vitro* experiments where shorter segments of BMV RNAs in the range of 500–2500 nts were mixed with CCMV capsid proteins, the resulting viruslike particles (VLPs) had a mixed population of pseudo $T = 2$ and $T = 3$ shells that were all overcharged [20]. RNA molecules shorter than 2000 nts were packaged in multiple copies, e.g., four in the case of 500 nt RNAs or two for 1000 nt RNAs in pseudo $T = 2$ capsids and two 1500 nt RNAs in $T = 3$ capsids.

Although the *in vitro* self-assembly studies show that RNA-based viruslike particles are overcharged, experiments with linear negatively charged polymers rather than virus RNAs are less conclusive. In fact, studies with linear polyanions, such as poly(styrene sulfonate) (PSS), have often focused attention more strongly on how the capsid size distribution is impacted upon by either the polymer length or the stoichiometry ratio of the capsid proteins and polymers [12,21]. What is known is that polymers, ranging in degree of polymerization from 1900 to 16 500 monomers, could all be encapsidated by a $T = 3$ structure, resulting in anything from a weakly to a highly overcharged structure [21]. *In vitro* self-assembly studies on mixtures of CCMV coat proteins and PSS chains as short as 180 monomers show a bimodal distribution of particle sizes corresponding to $T = 1$ and pseudo $T = 2$ structures [22]. According to these experiments there are on average two polymers in each $T = 1$ capsid (600 positive charges) and three in each $T = 2$ (1200 positive charges) [22]. Hence, the VLPs in this case are undercharged: The ratio of negative to positive charges is 0.6 for the $T = 1$ and 0.45 for $T = 2$ capsids. From all these findings it is not easy to extract a sound conclusion about the optimal length of the encapsulated polymer.

Several theoretical studies have shed light on the puzzling phenomenon of overcharging. Simulations on encapsulation of polymers with a fixed (quenched) level of branching as a model for RNA have shown that the level of overcharging is a sensitive function of the secondary and tertiary structures of the RNA [23,24]. Field-theoretic calculations presuming the branching to be annealed, not quenched, also have shown that the length of the encapsidated polymer, and hence the level of overcharging, increases as the number of branch points increases [25,26]. Although these theoretical studies confirm



(a) (b)

FIG. 1. (a) A $T = 1$ structure presented as $ISBF_{15,0}$, and (b) a $T = 3$ structure presented as $ISBF_{27,0}$.

that the topology of RNA is important to the phenomenon of overcharging, they also predict that the optimal number of charges on a linear polyelectrolyte must be *less* than the total number of charges on the inner capsid wall: These complexes must be undercharged rather than overcharged. This contrasts with the molecular dynamics (MD) simulations of Perlmutter *et al.*, which show that even linear polyanion encapsulation can lead to overcharging [23].

In virtually all theoretical studies focusing on the assembly of viral shells, the capsid has been assumed to be smooth and to have a uniform charge distribution in a region near the surface of the capsid [2,27–30]. However, as already alluded to, in most simple RNA viruses the positive charges reside on the RNA binding domains of the coat proteins, which are arranged according to the underlying icosahedral symmetry of the shell. This implies that the charge distribution must somehow reflect this icosahedral symmetry, certainly near the surface of the capsid, and perhaps less so away from it. Theoretically, the effects of localization of charge near the inner surface of the capsid on the encapsulation of the genome remain largely unexplored.

To remedy this, we study the impact of a nonuniform charge density on the optimal length of a genome encapsulated by small icosahedral viruses. Since $T = 1$ and $T = 3$ capsids have 60 and 180 RNA binding domains, respectively, we model capsids with 60 and 180 positively charged regions as shown in Fig. 1. We show how a nonuniform charge distribution, associated with the underlying icosahedral arrangement of the proteins part of a virus shell, results in a longer optimal genome length compared to a uniform charge distribution. This can give rise to the phenomenon of overcharging even for linear polyanions. The effects of a nonuniform charge distribution and the highly branched secondary structure of RNA, in particular for viral RNAs, conspire to greatly enhance overcharging. This allows for a larger amount of RNA to be packed in the same restricted interior of the virus shell, which arguably would be an evolutionary advantage for the virus.

Furthermore, we find that the optimal length of the genome, and as a result that of the number of encapsulated charges, depends on the detailed structure of RNA binding domains, i.e., the thickness, height, and charge density. This is consistent with the experimental findings of Ni *et al.* on the BMV in which mutations in the RNA binding domains that keep the number of charges constant but change their length and charge density impact upon the packaged RNA length [14]. These and many other experiments reveal the existence of intriguing results arising from the N-terminal domain topology [14,19,31,32]. A satisfactory theoretical approach needs to treat the coat protein topology (N-terminal domains), RNA folding, electrostatic interactions, and polymer confinement simultaneously. Our theoretical calculations allow us to single out the impact of length and charge density of the RNA binding domains without considering other effects, such as the impact of translational entropy and kinetic trapping, that make the interpretation of experiments and simulations difficult.

The paper is organized as follows. In the next section, we introduce the model and derive the equations that we will employ later. In Sec. III, we present our results corresponding to the capsid nonuniform charge distribution as well as RNA branching. Section IV discusses the impact on the capsid stability and overcharging phenomena of the length and charge density of N-terminal tails and the capsid radius. Finally, we present our conclusion and summarize our findings.

II. MODEL

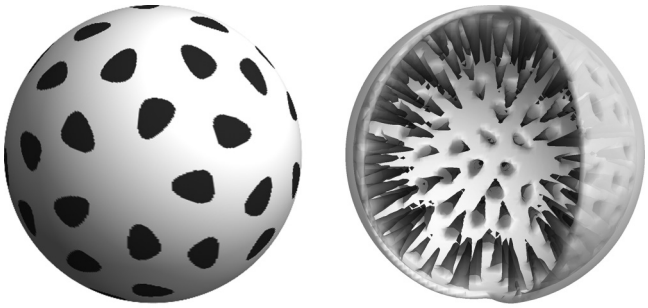
Our model consists of a mean-field theory that includes the entropic and steric contributions of the polyelectrolyte and the electrostatic interactions between the polyelectrolyte and the capsid. We initially model the genome as a flexible linear polyelectrolyte that interacts attractively with the positive charges residing on the binding domains and postpone the discussion of the impact of the RNA secondary structure to Sec. II A.

The free energy of a confined polyelectrolyte confined in a salt solution interacting with an external charge distribution can, within the ground-state approximation, be written as

$$\beta F = \int d^3\mathbf{r}\Big[\frac{1}{6}a^2|\nabla\Psi(\mathbf{r})|^2 + \frac{1}{2}\upsilon\Psi^4(\mathbf{r}) - \frac{1}{8\pi\lambda_B}|\nabla\beta e\Phi(\mathbf{r})|^2 - 2\mu\,\cosh[\beta e\Phi(\mathbf{r})] + \beta\tau\Psi^2(\mathbf{r})\Phi(\mathbf{r}) + \beta\rho_0(\mathbf{r})\Phi(\mathbf{r})\Big], \quad (1)$$

with β as the reciprocal temperature in units of energy, a as the statistical step or Kuhn length of the polymer, υ as the effective excluded volume per monomer, $\lambda_B = e^2\beta/4\pi\epsilon$ as the Bjerrum length, ϵ as the dielectric permittivity, e as the elementary charge, μ as the density of monovalent salt ions, and τ as the linear charge density of the chain. As usual, the dielectric permittivity of the medium is presumed constant [33].

The fields $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$ are the monomer density field and the electrostatic potential of the mean force, respectively. The positive charge density $\rho_0(\mathbf{r})$ is placed in an icosohedrally symmetric distribution either on the capsid surface as shown in Figs. 1(a) and 1(b) or extending into the interior of the capsid along the N-terminal tails as in Fig. 2(b). Extremizing the free energy with respect to the $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$ fields subject to the constraint that the total number of monomers inside the capsid is constant [34] $N = \int d^3\mathbf{r}\,\Psi^2(\mathbf{r})$ results in two self-consistent



(a) (b)

FIG. 2. The charge distributions from the capsids for (a) a $T = 1$ thin capsid. The black spots show the regions with a uniform surface charge density. The charges are smeared on the surface representing the thin capsid model. (b) A $T = 3$ thick capsid. The charges are extended into the interior of the capsid.

nonlinear field equations,

$$\frac{a^2}{6}\nabla^2\Psi = -\mathcal{E}\Psi(\mathbf{r}) + \beta\tau\Phi(\mathbf{r})\Psi(\mathbf{r}) + \upsilon\Psi^3, \tag{2a}$$

$$\frac{\beta e^2}{4\pi\lambda_B}\nabla^2\Phi(\mathbf{r}) = +2\mu e \sinh \beta e\Phi(\mathbf{r}) - \tau\Psi^2(\mathbf{r}) - \rho(\mathbf{r}), \tag{2b}$$

with $\mathcal{E}$ as the Lagrange multiplier enforcing the fixed number of monomers. Note $\rho(\mathbf{r})$ here is the volume charge density that will be set to zero if there are no charges extended to the interior of the capsid. The boundary conditions for the electrostatic potential inside and outside of the capsid that we model as a sphere of radius R are as follows:

$$\hat{n}\cdot\nabla\Phi_{\text{in}}|_{r=R} - \hat{n}\cdot\nabla\Phi_{\text{out}}|_{r=R} = 4\pi\lambda_B\sigma(\theta,\phi)/\beta e^2, \tag{3a}$$

$$\Phi_{\text{in}}(r)|_{r=R} = \Phi_{\text{out}}(r)|_{r=R}, \tag{3b}$$

$$\Phi_{\text{out}}(r)|_{r=\infty} = 0. \tag{3c}$$

with $\sigma(\theta,\phi)$ as the surface charge density. In the case of a space charge distribution $\rho \neq 0$, then we assume $\sigma = 0$. If the charges are localized to the surface, then $\sigma \neq 0$, but the volume charge density is $\rho = 0$. Thus, if the charges associated with the capsid are lying completely on the capsid wall, the volume charge density $\rho(\mathbf{r}) = 0$ in Eq. (2b) and the charge from the capsid are modeled as the surface charge $\sigma(\theta,\phi)$ in Eq. (3a). We discuss the exact forms of $\sigma(\theta,\phi)$ and $\rho(\mathbf{r})$ in Sec. II B.

We use Dirichlet $\Psi(r)|_{r=R} = 0$ boundary conditions for the chain density at the capsid wall, but our findings are robust, and we found the same results for the Neumann boundary condition $\partial_r\Psi(r)|_{r=R} = 0$. Although Eq. (2a) applies to a linear chain, a similar formalism can be employed to obtain the free energy of RNA modeled as a branched polymer trapped in a viral shell [25] as explained in the next section.

A. Branched polymer

To examine the combined effect of the secondary structure of RNA and nonuniform capsid charge distribution in this paper, we model RNA as an annealed branched polymer and add to Eq. (1) the following terms:

$$-\frac{1}{\sqrt{a^3}}\left(f_e\Psi + \frac{a^3}{6}f_b\Psi^3\right), \tag{4}$$

which describe the statistics of an annealed branched polymer [25,26,35–39] with f_e and f_b as the fugacities of the end and branch points, respectively [25]. The field equations [Eq. (2a)] become

$$\frac{a^2}{6}\nabla^2\Psi = -\mathcal{E}\Psi(\mathbf{r}) + \beta\tau\Phi(\mathbf{r})\Psi(\mathbf{r}) + \upsilon\Psi^3 - \frac{f_e}{2\sqrt{a^3}} - \frac{\sqrt{a^3}}{4}f_b\Psi^2. \tag{5}$$

In this formalism, the stem loops or hair pins in RNA are considered as end points. The number of end and branch points N_e and N_b of the polymer depend on the fugacities $N_e = -\beta f_e\frac{\partial F}{\partial f_e}$ and $N_b = -\beta f_b\frac{\partial F}{\partial f_b}$. We consider only the case of a single encapsulated polymer with no closed loops, and thus we have the following constraint: $N_e = N_b + 2$. The fugacity of branch points f_b determines the degree of branching.

B. Icosahedral symmetric based function

To explicitly model the charged N-terminal tails, we employ icosahedral symmetric based functions (ISBFs) for the $T = 1$ and $T = 3$ structures with 60 and 180 positively charged regions, respectively. These functions are real valued, complete, and orthogonal and can be written as a sum over spherical harmonics [40],

$$\text{ISBF}_{l,n}(\theta,\phi) = \sum_{m=-l}^{+l} b_{l,n,m}Y_{l,m}(\theta,\phi). \tag{6}$$

The ISBF functions are indexed by the integers l and n, where $l(l+1)$ is the azimuthal separation constant. $n \in \{0,1,\ldots,N_l - 1\}$ indexes the different ISBFs, and N_l denotes the number of linearly independent ISBFs that can be constructed for a given l. The weights $b_{l,n,m}$ can be computed for each l by comparing the expansion of the icosahedrally symmetric set of δ functions in both spherical harmonics and ISBFs.

The coefficients, $b_{l,n,m}$ given in Eq. (6) become nonzero only when m is a multiple of five, corresponding to fivefold symmetry of the icosahedral group. As a function of the associated Legendre function $P_m^l(x)$, ISBFs [40] can easily be written as

$$\text{ISBF}_{l,n}(\theta,\phi) = \begin{cases} \sum_{m=0}^{+l}\frac{2}{1+\delta_{m,0}}b_{l,n,m}N_{l,m}P_m^l(\cos\theta)\cos(m\phi), & l\text{: even}, \\ \sum_{m=1}^{+l}2b_{l,n,m}N_{l,m}P_m^l(\cos\theta)\sin(m\phi), & l\text{: odd}, \end{cases} \tag{7}$$

with $N_{l,m} = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}$.

The charge distributions for structures with $T = 1$ and $T = 3$ icosahedral symmetry are modeled by the ISBF with $(l = 15, n = 0)$ and $(l = 27, n = 0)$, respectively, shown in

TABLE I. The coefficients of $b_{l,n,m}$ for $\mathrm{ISBF}_{15,0}$ ($T = 1$) and $\mathrm{ISBF}_{27,0}$ ($T = 3$) structures, see Eq. (6) in the paper.

l	$b_{l,0,5}$	$b_{l,0,10}$	$b_{l,0,15}$	$b_{l,0,20}$	$b_{l,0,25}$
15	0.51653	0.39131	−0.28298		
27	0.44330	−0.23513	−0.02788	0.41768	−0.27011

Figs. 1(a) and 1(b). The values of $b_{l,n,m}$ for $T = 1$ and $T = 3$ structures are given in Table I.

Assuming that there are no charges in the regions between N-terminal tails [see Fig. 2(a)], we set charge density equal to zero if the magnitude of the ISBFs is lower than a certain cutoff value C. Thus, the distance between the charged regions depends on the cutoff, and since we fix the total charge of the capsid, the charge density of the N-terminal domain changes as a function of the cutoff. We consider both the *thin capsid model* where the charges are smeared on the surface of the spherical capsid in 60 or 180 positions as shown in Fig. 2(a) and the *thick capsid model* where the charges extend into the capsid as shown in Fig. 2(b). For the thick capsid model, we assumed that there are 60 ($T = 1$) or 180 ($T = 3$) "bumpy" charged regions extended inside the capsid. To this end, we shifted and truncated $\mathrm{ISBF}_{15,0}$ and $\mathrm{ISBF}_{27,0}$ such that the capsid surface protrudes in 60 or 180 positions presenting peptide tails [Fig. 2(b)] with charges uniformly distributed in the volume of the protruded regions.

III. RESULTS

We solved the coupled equations given in Eqs. (2) for Ψ and Φ fields, subject to the boundary conditions given in Eqs. (3) through the finite element method. The polymer density profiles Ψ^2 as a function of the distance from the center of the shell r are shown in Figs. 3(a) and 3(b) in three dimensions (3D) and one dimension (1D), respectively. As illustrated in the figure, the polymer density is higher at the N-terminal regions. Note that the density at the wall in the regions between N-terminal tails is lower than that in the N-terminal domains but still higher compared to the capsid center even though the capsid wall is not charged between the tails.

We find that the optimal genome length increases for a nonuniform charge distribution as compared to that where the charge distribution is uniform. In fact, the free energy in addition becomes deeper indicating a higher efficiency of genome encapsulation. Furthermore, we find that the optimal genome length increases if the cutoff C is increased and that the distance between the charged regions correspondingly increases. That is, as the charges on the capsid are distributed *more* nonuniformly, the optimal genome length increases. Nevertheless, for the thin capsid model, we have not been able to observe the phenomenon of overcharging with linear chains, i.e., the number of charges on the genome is always lower than those on the capsids for all the parameter values that we tested. This is not the case for the thick capsid model as explained below.

Figure 4 illustrates the encapsulation free energy as a function of genome length for a $T = 3$ structure with the radius

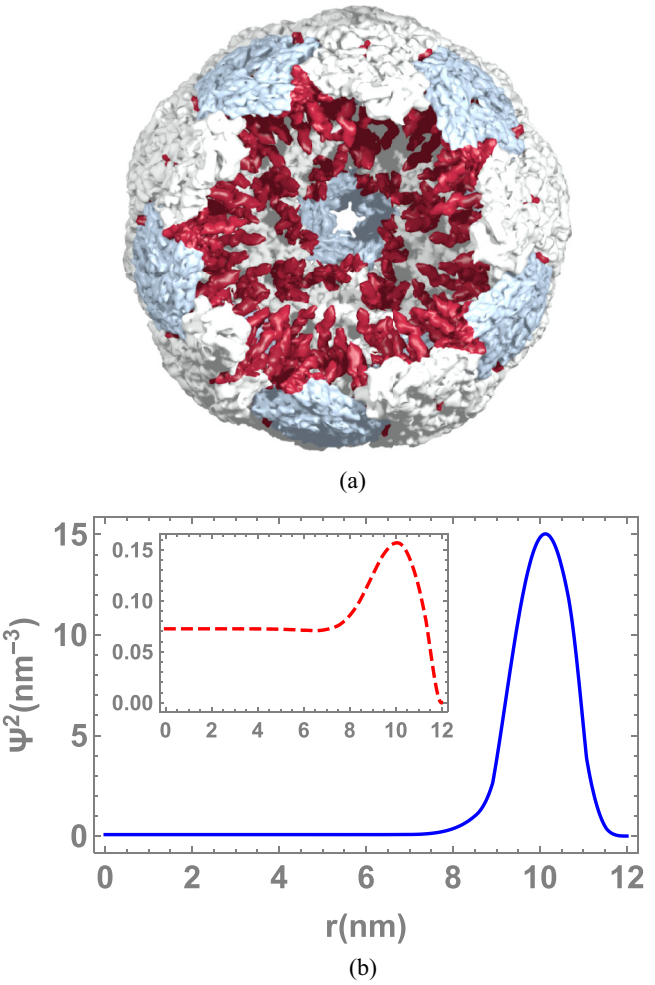


FIG. 3. Genome density profile of a $T = 3$ capsid in (a) a 3D view. The protruded regions represent RNA (red). The density of RNA between N-terminals is very small and not shown in the figure. (b) A 1D view as the function of capsid radius with nonuniform charge distribution. The figure shows the profiles along two different directions. The solid line corresponds to the direction in which the N-terminal tail is located, and the dashed line corresponds to the direction without the N-terminal tail (the inset graph). In the absence of surface charge density and the N-terminal tail (dashed curve), the density is still maximum close to the wall. The polymer is branched with $f_b = 3$, total monomer number = 2411, salt concentration $\mu =$ 100 mM, $R = 12$ nm, and $Qc = 1800$.

of capsid $R = 12$ nm at $\mu = 100$ mM salt concentrations for the thick capsid model [Fig. 2(b)]. The total number of charges is assumed to be $Q_c = 1800$ for a $T = 3$ structure, ten charges on each N-terminal tail. The dashed lines in Fig. 4 correspond to a capsid with a uniform charge density, and the dotted lines correspond to a nonuniform charge density. The lines with the shortest distance between the dashed lines and the dots correspond to that of a linear polymer. As illustrated in the figure, the minimum of the free energy moves towards longer chains if the charge distribution is nonuniform.

Figure 4 also shows the impact of RNA secondary structures on the optimal length of the encapsulated genome. The graphs in Fig. 4 correspond to $f_b = 0$ for a linear polymer and $f_b = 1.0$ and $f_b = 3.0$ for branched ones. The polymer

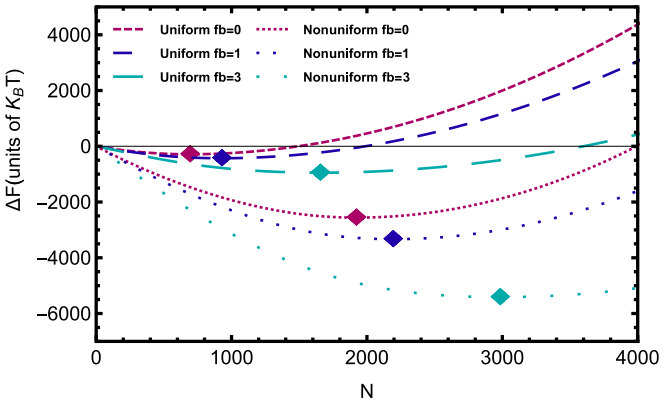


FIG. 4. Encapsulation free energy for a linear and branched polyelectrolyte as a function of monomer number for a capsid with uniform (dashed lines) and nonuniform (dotted lines) charge densities. For a linear chain the branching fugacity $f_b = 0$ and increases to $f_b = 1.0$ and $f_b = 3.0$ as the chain becomes more branched. The diamonds indicate the minimum of free energy. The other parameters used correspond to a $T = 3$ virus: total capsid charges on capsid $Q_c = 1800$, $a = 1.0$ nm, $v = 0.01$ nm^3 $\mu = 100$ mM, $R = 12$ nm, and tail length $= 4$ nm.

becomes more strongly branched as f_b increases. Note that in the figure the distance between dots or dashed lines increases as the fugacity or the number of branch points increases. The figure reveals that, as the degree of branching increases, the length of encapsulated genome increases for a capsid with a uniform charge density. This effect becomes stronger if we consider a nonuniform charge distribution. The diamonds in the figure indicate the optimal length of the genome. The ratios of the optimal length or the number of charges on RNA to the capsid total charge $Q_c = 1800$ from left to right in the figure are 0.39, 0.52, 0.92, 1.07, 1.22, and 1.66, which clearly shows a transition from undercharging towards overcharging. We note that we find the same behavior when employing a $T = 1$ instead of a $T = 3$ capsid.

IV. DISCUSSION AND SUMMARY

The reason for overcharging associated with the nonuniform charge distributions is twofold. A nonuniform charge distribution on the capsids obviously promotes a nonuniform genome density distribution. However, in order to have a more uniform polymer distribution with lower entropy cost, longer chains are preferably encapsulated to make the genome distribution more uniform in the regions between the N-terminal tails. Figure 5(a) illustrates this effect for a $T = 3$ structure with 180 tails as a plot of the optimal length of genome vs the capsid charge density. Note that, since the total charge of capsid is fixed, as we increase the charge density, we lower the volume of the N-terminal regions, which also is shown in the axis on the top of the graph. The vertical axis on the right-hand side of the figure shows the degree of overcharging. The circles in the figure correspond to a $\mu = 100$ mM salt concentration, and the squares correspond to $\mu = 500$ mM. For the hollow symbols the radii of the capsids are $R = 9.5$ nm, but for the solid symbols $R = 11.5$ nm. As shown in the figure, if we increase the area between

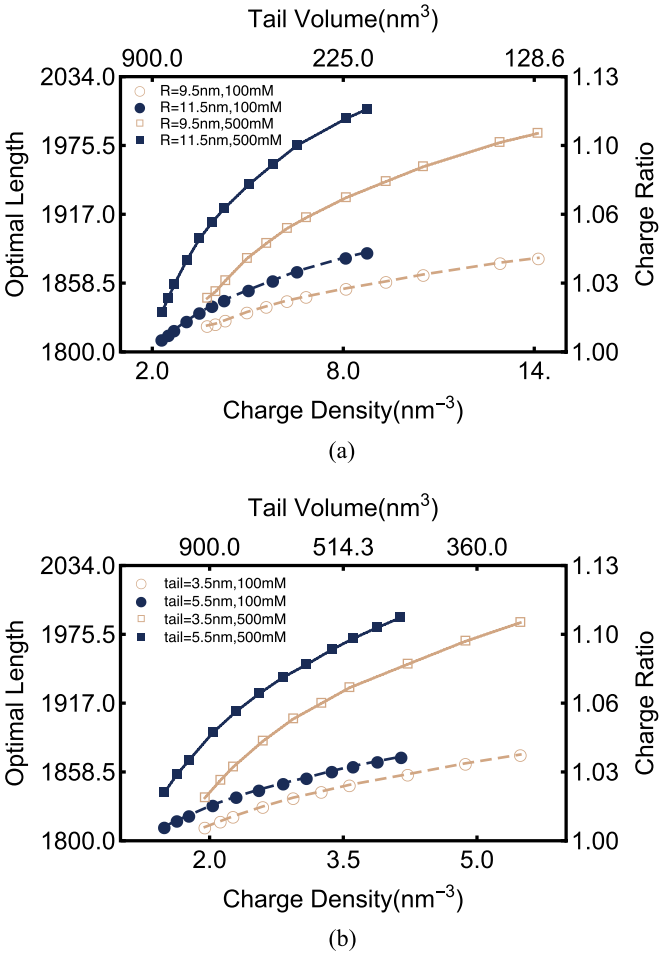


FIG. 5. Optimal genome length or charge ratio vs N-terminal charge density or volume occupied for $T = 3$. (a) The hollow symbols correspond to capsid radii of $R = 9.5$ nm, and the solid ones correspond to $R = 11.5$ nm with the N-terminal tail length 3.5 nm; (b) The hollow symbols correspond to a tail length of 3.5 nm, and the solid ones correspond to 5.5 nm for $R = 12.5$ nm. The other parameters are total charge $Qc = 1800$, salt concentration $\mu = 100$ mM (dashed lines), and $\mu = 500$ mM (solid lines).

the N-terminals or the radius of the capsid, the amount of overcharging increases at a given salt concentration.

However, the noted entropy effect cannot explain all the observations. At the physiological salt concentration of $\mu = 100$ mM, the genome only interacts with the capsid if it is sitting in vicinity of the capsid coat protein charges. This is due to the rather short range of electrostatic interactions at that salt concentration. The presence of N-terminals increases the region with which the genome interacts attractively through electrostatic interactions. Thus, the higher the salt concentration, the more important becomes the role of N-terminals. The figure shows that the overcharging is more pronounced at $\mu = 500$ mM. Also, the higher the salt concentration, the lower the electrostatic self-repulsion between genome monomers, which helps the encapsidation of longer chains.

We also examined the impact of the length of N-terminal domains in the thick capsid model, which corresponds to how far the charged regions extend into the interior of the capsid.

As illustrated in Fig. 5(b) for $T = 3$ capsids, more genome is encapsidated for longer N-terminal tails, which is again due to a larger interacting region for a fixed total number of charges on the capsid. The effect becomes more pronounced for higher salt concentrations as illustrated in the figure.

To summarize, we have studied the phenomena of overcharging observed in many viruses. Previous mean-field theories as well as the experimental studies of CCMV capsid proteins with short linear polymers have indicated the resulting VLPs are undercharged [25,26,41–44]. However, MD simulations revealed overcharging can happen even for linear polymers, and the question is why [23]. In this paper, we showed that the nonuniform charge distribution increases both the stability and the amount of genome that can be assembled by CPs as a result of what in essence is entropy. For a thin capsid model with the charges smeared flatly on the surface, longer chains are encapsulated, but we have not been able to observe overcharging with linear polymers. This indicates that overcharging for linear systems is primarily due to the charged N-terminal regions that protrude into the interior of the capsid. The N-terminal regions increase the regions in which the genome can interact with the capsid proteins and thus result in the encapsidation of longer chains. This latter effect is stronger at higher salt concentrations. We find that the combined effect of RNA base pairing, which gives rise to the genome branching, and nonuniform charge distribution can explain the pronounced charge inversion observed in viruses.

ACKNOWLEDGMENTS

The authors would like to thank V. Lorman for useful discussions. This work was supported by the National Science Foundation through Grant No. DMR-1310687 (R.Z.).

[1] J. B. Bancroft, Adv. Virus Res. **16**, 99 (1970).
[2] R. F. Bruinsma, Eur. Phys. J. E **19**, 303 (2006).
[3] S. N. Fejer, D. Chakrabarti, and D. J. Wales, Nano. Lett. **4**, 219 (2010).
[4] D. C. Rapaport, Phys. Rev. E **70**, 051905 (2004).
[5] J. Wagner and R. Zandi, Biophys. J. **109**, 956 (2015).
[6] T. Chen, Z. Zhang, and S. C. Glotzer, Proc. Natl. Acad. Sci. USA **104**, 717 (2007).
[7] S. Paquay, H. Kusumaatmaja, D. J. Wales, R. Zandi, and P. van der Schoot, Soft Matter **12**, 5708 (2016).
[8] J. Sun, C. DuFort, M.-C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, and B. Dragnea, Proc. Natl. Acad. Sci. USA **104**, 1354 (2007).
[9] R. Kusters, H.-K. Lin, R. Zandi, I. Tsvetkova, B. Dragnea, and P. van der Schoot, J. Phys. Chem. B **119**, 1869 (2015).
[10] M. F. Hagan and R. Zandi, Curr. Opin. Virol. **18**, 36 (2016).
[11] H. Tao, Z. Rui, and B. I. Shklovskii, Physica A **387**, 2495 (2008).
[12] F. D. Sikkema, M. Comellas-Aragones, R. G. Fokkink, B. J. M. Verduin, J. Cornelissen, and R. J. M. Nolte, Org. Biomol. Chem. **5**, 54 (2007).
[13] Y. P. Ren, S. M. Wong, and L. Y. Lim, J. Gen. Virol. **87**, 2749 (2006).
[14] P. Ni, Z. Wang, X. Ma, N. C. Das, P. Sokol, W. Chiu, B. Dragnea, M. Hagan, and C. C. Kao, J. Mol. Biol. **419**, 284 (2012).
[15] A. Losdorfer Bozic, A. Siber, and R. Podgornik, J. Biol. Phys. **39**, 215 (2013).
[16] M. C. Daniel, I. B. Tsvetkova, Z. T. Quinkert, A. Murali, M. De, V. M. Rotello, C. C. Kao, and B. Dragnea, ACS Nano **4**, 3853 (2010).
[17] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, Virology **277**, 450 (2000).
[18] H.-K. Lin, P. van der Schoot, and R. Zandi, Phys. Biol. **9**, 066004 (2012).
[19] V. Sivanandam, D. Mathews, R. Garmann, G. Erdemci-Tandogan, R. Zandi, and A. L. N. Rao, Sci. Rep. **6**, 26328 (2016).
[20] M. Comas-Garcia, R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart, J. Virol. **86**, 12271 (2012).
[21] Y. F. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, Biophys. J. **94**, 1428 (2008).
[22] R. D. Cadena-Nava, Y. F. Hu, R. F. Garmann, B. Ng, A. N. Zelikin, C. M. Knobler, and W. M. Gelbart, J. Phys. Chem. B **115**, 2386 (2011).
[23] J. D. Perlmutter, C. Qiao, and M. F. Hagan, eLife **2**, e00632 (2013).
[24] S. W. Singaram, R. F. Garmann, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, Acc. Chem. Res. **119**, 13991 (2015).
[25] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, Phys. Rev. E **89**, 032707 (2014).
[26] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, Phys. Rev. E **94**, 022408 (2016).
[27] R. Zandi and P. van der Schoot, Biophys. J. **96**, 9 (2009).
[28] A. L. Bozic, A. Siber, and R. Podgornik, J. Biol. Phys. **38**, 657 (2012).
[29] P. van der Schoot and R. Bruinsma, Phys. Rev. E **71**, 061928 (2005).
[30] P. van der Schoot and R. Zandi, J. Biol. Phys. **39**, 289 (2013).
[31] D. Marshall and A. Schneemann, Virology **285**, 165 (2001).
[32] S. B. Larson, R. W. Lucas, and A. McPherson, J. Mol. Biol. **346**, 815 (2005).
[33] M. Janssen, A. Härtel, and R. van Roij, Phys. Rev. Lett. **113**, 268501 (2014).
[34] H. Ji and D. Hone, Macromolecules **21**, 2600 (1988).
[35] T. C. Lubensky and J. Isaacson, Phys. Rev. A **20**, 2130 (1979).
[36] T. T. Nguyen and R. F. Bruinsma, Phys. Rev. Lett. **97**, 108102 (2006).
[37] S. I. Lee and T. T. Nguyen, Phys. Rev. Lett. **100**, 198102 (2008).
[38] K. Elleuch, F. Lequeux, and P. Pfeuty, J. Phys. I (France) **5**, 465 (1995).
[39] J. Wagner, G. Erdemci-Tandogan, and R. Zandi, J. Phys.: Condens. Matter **27**, 495101 (2015).

[40] Y. Zheng and P. C. Doerschuk, SIAM J. Math. Anal. **32**, 538 (2000).
[41] A. Šiber and R. Podgornik, Phys. Rev. E **76**, 061906 (2007).
[42] A. Šiber and R. Podgornik, Phys. Rev. E **78**, 051915 (2008).
[43] C. L. Ting, J. Z. Wu, and Z. G. Wang, Proc. Natl. Acad. Sci. USA **108**, 16986 (2011).
[44] A. Šiber, R. Zandi, and R. Podgornik, Phys. Rev. E **81**, 051919 (2010).