RNA topology remoulds electrostatic stabilization of viruses

Gonca Erdemci-Tandogan,¹ Jef Wagner,¹ Paul van der Schoot,²,³ Rudolf Podgornik,⁴,⁵ and Roya Zandi¹

¹Department of Physics and Astronomy, University of California, Riverside, California 92521, USA
²Group Theory of Polymers and Soft Matter, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
³Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands
⁴Department of Theoretical Physics, J. Stefan Institute, SI-1000 Ljubljana, Slovenia
⁵Department of Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia

Simple RNA viruses efficiently encapsulate their genome into a nano-sized protein shell—the capsid. Spontaneous co-assembly of the genome and the capsid proteins is driven predominantly by electrostatic interactions between the negatively charged RNA and the positively charged inner capsid wall. Using field theoretic formulation we show that the inherently branched RNA secondary structure allows viruses to maximize the amount of encapsulated genome and make assembly more efficient, allowing viral RNAs to out-compete cellular RNAs during replication in infected host cells.

Simple viruses encapsulate their genetic material into a protein shell, measuring no more than about 15 nm across for the smallest of viruses and about 28 nm for a typical (plant) virus ¹ ². Under many circumstances, in vitro virus assembly is spontaneous and driven primarily by electrostatic interactions between negative charges on the backbone of the polynucleotide, usually single-stranded (ss) RNA, and positive charges on the virus coat proteins ³ ¹⁰. However, recent experiments indicate that RNA plays a role that goes beyond its polyelectrolyte nature, as some RNAs are encapsulated more efficiently than others ¹¹. For example, when viral RNA1 of BMV (brome mosaic virus) and CCMV (cowpea chlorotic mosaic virus) are mixed in solution with the capsid proteins from CCMV, the BMV RNA is packaged 3 times more efficiently ¹¹, 12. As the two RNAs differ only in the amount of branching and their tertiary structure, both a straightforward consequence of their template nucleotide sequence, there must be a tight connection between capsid packing preferences and the structure of RNA ¹² ¹³.

In many viruses the number of negative charges on the ssRNA is larger than the number of positive charges on the virus coat proteins ¹⁴ ¹⁷. Belyi and Muthukumar examined a sample of actual viruses and found that the ratio of the RNA charge to structural charge on the inner capsid surface to be around 1.6 ¹⁰. While, it seems to be feasible to encapsulate linear polymers with number of charges as much as nine times that on the capsid proteins ¹⁸, which would result in strong “overcharging” of the virion, recent experiments show that the optimal number of charges residing on the linear polyelectrolyte (PE) is less than the total number of charges on the inner surface of viral shells ¹⁹ implying “undercharging” of the complex of capsid proteins with linear polyelectrolytes. This naturally leads to the question of which RNA features are implicated in the overcharging of the virion. In what follows we show that RNA secondary structural features, such as branching, have a pronounced effect on the genome encapsulation capacity and even drive the selective preference for viral RNAs compared to the cellular RNAs, as it is shown that the latter has a more branched structure.

In virtually all theoretical studies published to date, investigating the impact of electrostatic interactions on the structure and stability of simple viruses, the genome is modeled as a simple linear polyelectrolyte chain ¹⁵ ²²–²⁴. Nevertheless, recent experiments alluded to above, reveal the importance of RNA structure that goes beyond its polyelectrolyte nature as a linear charged chain ¹¹, ¹², ¹⁷, ²², ²⁰. Intra-chain base paring, promoted by hydrogen bonding between mutually complementary nucleotides along the backbone, leads to a highly branched structure of the RNA molecule that furthermore pro-

FIG. 1: Encapsulation free energy as a function of monomer numbers for a linear polymer with \( f_b=0 \) (solid and dotted lines) and a branched polymer with \( f_b=3 \) (dashed and dotted-dashed lines) at two different values of \( \mu \), corresponding to salt concentrations 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines). The arrow indicates the monomer number at which the full virus particle (capsid + polyelectrolyte) becomes neutral. Inset shows the position of the minimum \( N_{\text{min}} \) vs the branching fugacity \( f_b \) for \( \mu=100 \) mM salt concentration. Other parameters used are \( v=0.5 \), \( \tau=-1 \), \( \sigma=0.4 \), \( b=12 \), and \( T=300 K \), typical for RNA and virus capsids ² ¹³ ²⁰ ²¹.
The inset to Fig. 1, showing the position of the minimum towards longer chains as the branching fugacity increases. The results are evident from Fig. 1 showing the displacement in the position of the encapsulation free energy minimum versus longer chains as the branching fugacity increases. The inset to Fig. 1 showing the position of the minimum $N_{min}$ as a function of the propensity for branching $f_b$, directly demonstrates this effect.

The figure also illustrates that the encapsulation free energy becomes more negative with increasing propensity of RNA to form branch points for a given number of monomers. This stabilization behavior suggests that branching is not only conducive to more efficient packing of the genome material into the virus shell, but also allows viral RNAs that have more branch points than other types of cellular RNAs [12], to out-compete the latter during replication in infected, susceptible host cells.

To obtain the optimal length and the free energy associated with the encapsulated RNA inside a capsid, we model RNA as a generic flexible branched polyelectrolyte. Because of the physical character of the base pairing, the degree of branching of RNAs is statistical and may in the process of encapsulation be affected by interaction with the charges located on or near the inner surface of the protein coat. To this end, we consider only the case of annealed branched polymers in this letter.

Further, we consider that the RNA interacts with positive charges residing on the inner surface of a sphere, where for simplicity we additionally presume that the charges are not localized but smeared out uniformly. For a large proportion of viruses the positive charges are indeed located on the inner surface of the capsid that in reality is not a perfect sphere but has structure on the nanometer scale [21]. For some viruses positively charged disordered domains on the coat proteins point into the capsid cavity in a brush-like fashion [14], a feature that we do not include in our coarse-grained model at this stage.

In the mean-field, ground-state approximation, the free energy of a negatively charged polymeric chain confined to a positively charged, infinitely thin spherical shell can be written as

$$ F = \int d^3r \left[ \frac{1}{2} |\nabla \Psi(r)|^2 + W[\Psi(r)] \right. $$

$$ \left. - \frac{1}{8\pi\varepsilon_0} |\nabla \Phi(r)|^2 - 2\mu \cosh \left[ \Phi(r) \right] + \tau \Phi(r) \Psi^2(r) \right] + \int d^2r \left[ \sigma \Phi(r) \right]. $$

All quantities in Eq. (1) are dimensionless, so energies are in units of thermal energy $k_BT$ and lengths in units of the statistical step length (Kuhn length) of the polymer, $a$. Here, $\tau$ denotes the linear charge density of the polymer, $\sigma$ the surface charge density of the shell, $\Psi^2(r)$ the monomer density at position $r$, and $\Phi(r)$ the mean electrostatic potential. The parameter $\mu$ is the fugacity (density) of the monovalent salt ions, and corresponds to the concentration of salt ions in the bulk. The (dimensionless) Bjerrum length, $\lambda_B$, is a measure of the dielectric constant of the solvent, corresponding to about 0.7 nm for water at room temperature. The square gradient term in the first line of Eq. (1) describes the entropic cost for a non-uniform polymer density. The $W[\Psi]$ term describes the statistics of an annealed branched polymer [27–30], given explicitly by

$$ W[\Psi] = \frac{1}{2} \nu \Psi^4 - f_e \Psi - \frac{f_b}{6} \Psi^3, $$

where $\nu$ is the (dimensionless) excluded volume and $f_e$ and $f_b$ are the fugacities of the end-and branch-points, respectively. The last two lines in Eq. (1) describe the electrostatic interactions between the polyelectrolyte, the salt ions, and the charged capsid at the level of Poisson-Boltzmann theory [31].

In our description, the stem-loop or hairpin configurations in RNA structures are counted as end points. The number of end and branch points $N_e$ and $N_b$ of the polymer depend on the fugacities $f_e$ and $f_b$ through

$$ N_e = -f_e \frac{\partial F}{\partial f_e} \quad \text{and} \quad N_b = -f_b \frac{\partial F}{\partial f_b}. $$

Since we consider only the case of a single encapsulated polymer with no closed loops, there is a constraint on the number of end and branch points,

$$ N_e = N_b + 2, $$

with the degree of branching controlled by the fugacity of branch points $f_b$. The chain is linear if $f_b = 0$, and becomes more branched as $f_b$ increases. The fugacity of endpoints $f_e$ is set through the above constraint, Eq. (4). In addition, the total number of polyelectrolyte monomers inside the capsid is fixed [32], i.e.,

$$ N = \int d^3r \Psi^2(r), $$

which we enforce by introducing a Lagrange multiplier, $E$, when minimizing the free energy. We obtain the polyelectrolyte profile and electrostatic potential by varying the free energy functional with respect to fields $\Psi(r)$ and $\Phi(r)$. The resulting coupled set of non-linear equations describes the monomer density field, $\Psi$, and the electrostatic potential $\Phi_{in}$, in interior of the capsid, and the usual Poisson-Boltzmann equation for the electrostatic potential, $\Phi_{out}$, in the exterior of the
capsid, viz.,

\[
\begin{align*}
\frac{1}{6} \nabla^2 \Psi(r) &= -E\Phi_{\text{in}}(r) + \tau \Phi_{\text{out}}(r) + \frac{1}{2 \partial \Psi} \\ \nabla^2 \Phi_{\text{in}}(r) &= \frac{1}{\lambda D} \sinh \left[ \Phi_{\text{in}}(r) \right] - \frac{\tau}{2 \lambda^2 D \mu} \Psi^2(r) \\ \nabla^2 \Phi_{\text{out}}(r) &= \frac{1}{\lambda D} \sinh \left[ \Phi_{\text{out}}(r) \right]
\end{align*}
\]

where \( \lambda_D = 1/\sqrt{8\pi \lambda B \mu} \) is the (dimensionless) Debye screening length. The polymer segment concentration outside the capsid is assumed to be zero, \( \Psi = 0 \). Eqs. (6) represent a set of coupled nonlinear differential equations that can only be solved numerically, subject to appropriate boundary conditions.

Assuming that the positive surface charge density, \( \sigma \), is fixed, the electrostatic boundary condition is obtained by minimizing the free energy with respect to \( \Phi \) on the surface, \( \hat{n} \cdot \nabla \Phi_{\text{in}} - \hat{n} \cdot \nabla \Phi_{\text{out}} = 4\pi \lambda_B \sigma \). The choice of boundary conditions for \( \Psi \) depends on how the polymer interacts with the capsid surface through non-electrostatic forces. The strong short-ranged repulsion (as would be the case if we had included an excluded volume term between the polyelectrolyte monomers and the capsid proteins) leads to Dirichlet boundary conditions. However, it turns out that for the large surface charge densities relevant to viruses, our conclusions are robust and do not depend on the choice of boundary condition. Here, we present our results only for Neumann boundary conditions, resulting naturally from the minimization of the free energy.

The overall dimensionless monomer density profiles \( C(r) = \Psi(r)^2 \) as a function \( r = |r| \) the distance from the center of the cavity are shown in Fig. 2 for a linear polymer with \( f_b = 0 \) and a branched polymer with \( f_b = 3.0 \) of an equal number of segments, \( N = 1000 \), enclosed in a spherical shell. The radius of the capsid is taken to be \( b = 12 \) in units of the polymer Kuhn length that for our purpose is of the order of 1 nm [20]. Both types of polymers can adsorb onto the surface and, interestingly, the branched polymer is adsorbed more densely onto the surface than the linear chain.

We also investigated the spatial inhomogeneity in our annealed branched polymer model of RNA. In the inset (a) of Fig. 2 we plot the dimensionless density of endpoints \( C_e(r) = f_e \Phi(r) \) (solid line) and branches \( C_b(r) = f_b \Phi^3(r) \) (dashed lines), obtained from Eq. (3). The inset (b) of Fig. 2 illustrates the fractions of endpoints \( C_e/C \) (solid line) and fraction of branches \( C_b/C \) (dashed lines) as a function of the distance from the center of the capsid. We observe that most branching takes place within a two Debye length layer, thus very near the capsid wall where the concentration of segments is high and the local gradient in density is the largest. This is straightforward to understand as branching increases the local density allowing more segments to interact with the wall. Figure 3 also shows that end points are dominantly distributed over the capsid interior. Thus branching can affect not only the segment distribution but also the structure of the adsorbed layer making both quite non-uniform.

If we insert the equilibrium profiles for \( \Psi \) and \( \Phi \) obtained from Eqs. (6) for a fixed Lagrange multiplier \( E \) and branch point fugacity \( f_b \) into Eq. (4) and Eq. (5), we obtain the free energy of the chain-capsid complex, \( F \), and the monomer number, \( N \). By varying the Lagrange multiplier \( E \), we are then able to plot \( F \) vs \( N \) as shown in Fig. 4, which confirms that the free energy minimum moves towards longer chains, i.e., allowing more monomers to be encapsulated into the viral shell. In the inset to Fig. 4 we plot the optimal polymer length \( N_{\text{min}} \).
(defined as the position of the free energy minimum) versus the branch point fugacity $f_b$ for $\mu$, corresponding to 100 mM. As illustrated in Fig. 1, this effect is more pronounced at high salt concentrations. This is mainly due to the fact that high salt concentration screens electrostatic interactions, and thus intrinsic features of RNA structure such as branching will become more relevant. For low salt concentration, electrostatics overwhelms all the other interactions and branching becomes less important. For example, as shown in Fig. 1 for 100 mM salt, the optimal number of monomers for the linear polymer is about 534, but for the branched polymer it increases by more than two times, becoming as large as 1211. The arrow on the $N$ axis of Fig. 1 corresponds to $N = 4\pi b^2 \sigma$ representing a neutral system where the number of positive charges on the capsid wall is equal to the number of negative charges on the chain.

This observation reveals that overcharging in viral particles could be a direct result of the secondary, i.e., branched, structure of viral RNA. We note here that we repeated the above calculations for different surface charge densities, relevant to different viral capsids ($0.3 \leq \sigma \leq 0.9$), and found that for all cases, the number of charges on linear polymers is less than the number of positive charges on the capsid. Quite interestingly, we also found that for any given linear charge density of the chain, the optimal length of encapsulated branched polymers is always larger than that of linear polymers.

Figure 1 also reveals the second important point, viz., that the free energies associated with branched polymers have deeper minima than those for linear polymers for a set of salt concentrations. This effect explains why some RNAs are encapsulated more efficiently than other RNAs, or other linear polyelectrolytes for that matter.

The increased packaging efficiency is further illustrated by comparing the osmotic pressure as seen in Fig. 4 resulting from the encapsulation of a linear polyelectrolyte with that of a branched one, as defined in Ref. [31],

$$P(N) = \frac{1}{4\pi\sigma} (\frac{\partial F}{\partial Q} |_{Q=N} - \frac{\partial F}{\partial Q} |_{Q=N=0})$$

The osmotic pressure quantifies the force exerted on the virus capsid by the genome. Figure 4 shows that osmotic pressure for our system has negative values, as expected for all systems that spontaneously self-assemble, and furthermore that the magnitude of osmotic pressure decreases as the salt concentration increases, revealing the important stabilizing role of electrostatic interaction in virus assembly.

If coupling between RNA branching and electrostatics represents a robust mechanism, the details of its description should not be qualitatively important. To this end, it is interesting to compare our results for encapsulated charged branched polymer based on a field theoretic Ansatz for the statistics of branched polymers with a very simple model with short-ranged attractive interaction between different chain segments mimicking the self-pairing of RNA bases [33]. We consider a linear polymer and now define the $W[\Psi]$ term as

$$W[\Psi] = \frac{1}{2} (s - w) \Psi^4 + u \Psi^6,$$

with $s$ the average fraction of base-pairs and $w$ the binding energy. We also include the next term in the virial expansion in order to stabilize the free energy since the total coefficient in front of the $\psi^4$ term can become negative. Calculating $F$ vs. $N$ curves for increasing values of $s$, the average fraction of self-paired bases, we find the same qualitative behavior as for increasing branching fugacity: the position of the minimum moves towards longer polymers (larger $N$) and the depth of the minimum increases for increasing $s$. In addition, as was the case for the branching model, the effect is stronger at higher salt concentrations.

Both models described above show that the total charge of the genome inside the capsid is larger than the one residing on the capsid interior and that the virion is thus overcharged. Our analysis clearly shows that the genomic function of RNA, as encoded in its sequence that engenders its branched secondary structure, plays an important role in the self-assembly of ssRNA viruses. The branched secondary structure of RNA, treated with either branching or self-pairing models, promote overcharging of the virion and stabilize its equilibrium configuration.

In order to explain the experiments noted in the introduction on the competition between RNA of CCMV and BMV through the theory presented above, we calculated the number of branch points for both RNAs. In particular, we used RNASubopt, a program in the Vienna RNA package [34], to generate an ensemble of secondary structures for genome sequences of RNA1 of BMV and CCMV. We then calculated the thermally averaged number of branch points from the secondary structures of each RNA. We found that RNA1 of BMV has higher average number of branch points (65) than CCMV (60.5) confirming that in the absence of specific interactions RNA1 of BMV would be preferentially packaged over RNA1 of CCMV, consistent with the experimental results of Comas-Garcia et al. [11, 37]. While one has to be cautious about the results of RNA packages for longer sequences, the Vienna RNA Package [34] has been used to calculate thermally averaged properties of viral genomes with lengths of 2500-7000 nt, and important results have been obtained [12].

A comprehensive investigation of the physico-chemical parameters that impact capsid formation could have great potential in the development of anti-viral therapies and a systematic understanding of the processes involved in viral infection.

The authors would like to thank Mehran Kardar and Aaron Yoffe for useful discussions. This work was supported by the National Science Foundation through Grant No. DMR-06-45668 (R.Z.).
[1] J. B. Bancroft, Adv. Virus Res. 16, 99 (1970).
[2] R. F. Bruinsma, Euro. Phys. J. E 19, 303 (2006).
[3] 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).
[4] Y. P. Ren, S. M. Wong, and L. Y. Lim, J. Gen. Virol. 87, 2749 (2006).
[5] R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart, J. Phys. Chem. B 115, 2386 (2011).
[6] R. Zandi and P. van der Schoot, Biophys. J. 96, 9 (2009).
[7] A. L. Bozic, A. Siber, and R. Podgornik, J. Biol. Phys. 38, 657 (2012).
[8] A. Siber, R. Zandi, and R. Podgornik, Phys. Rev. E 81, 051919 (2010).
[9] O. M. Elrad and M. F. Hagan, Phys. Biol. 7, 045003 (2010).
[10] M. F. Hagan, Adv. Chem. Phys. 155, (in press) (2013).
[11] A. Siber, R. Zandi, and R. Podgornik, Phys. Rev. E 81, 051919 (2010).
[12] A. Siber and R. Podgornik, Phys. Rev. E 78, 051915 (2008).
[13] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, Monatsh. Chem. 125, 167 (1994).
[14] The standard deviation for the number of branch points in the ensemble of RNA secondary structures is around 2.5 for the RNA1 sequence of both BMV and CCMV.