

# The effect of RNA stiffness on the self-assembly of virus particles

**Siyu Li<sup>1</sup> , Gonca Erdemci-Tandogan<sup>2</sup>, Paul van der Schoot<sup>3,4</sup> and Roya Zandi<sup>1</sup>**

<sup>1</sup> Department of Physics and Astronomy, University of California, Riverside, CA 92521, United States of America

<sup>2</sup> Department of Physics, Syracuse University, Syracuse, NY 13244, United States of America

<sup>3</sup> Group Theory of Polymers and Soft Matter, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, Netherlands

<sup>4</sup> Institute for Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, Netherlands

E-mail: [sli032@ucr.edu](mailto:sli032@ucr.edu)

Received 31 October 2017, revised 6 December 2017

Accepted for publication 13 December 2017

Published 27 December 2017



## Abstract

Under many *in vitro* conditions, some small viruses spontaneously encapsidate a single stranded (ss) RNA into a protein shell called the capsid. While viral RNAs are found to be compact and highly branched because of long distance base-pairing between nucleotides, recent experiments reveal that in a head-to-head competition between an ssRNA with no secondary or higher order structure and a viral RNA, the capsid proteins preferentially encapsulate the linear polymer! In this paper, we study the impact of genome stiffness on the encapsidation free energy of the complex of RNA and capsid proteins. We show that an increase in effective chain stiffness because of base-pairing could be the reason why under certain conditions linear chains have an advantage over branched chains when it comes to encapsidation efficiency. While branching makes the genome more compact, RNA base-pairing increases the effective Kuhn length of the RNA molecule, which could result in an increase of the free energy of RNA confinement, that is, the work required to encapsidate RNA, and thus less efficient packaging.

Keywords: self-assembly, virus, branched polymer, polymer field theory

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Ribonucleic acid (RNA) is one of the molecules of life, which plays a central role in the cell as information carriers, enzymes, gene regulators, etc. It is made out of four elementary building nucleotides, being A(denine), G(uanine), C(ytosine) and U(ricil) [1]. As shown by Crick and Watson, purines (A,G) pair with complementary pyrimidines (C,U), leading primarily to the pairs CG and AU. There exist also so-called wobble pairs of GU. Single stranded RNA is quite flexible with a Kuhn length of, depending on the ionic strength of the solution, one or two nm [2], and can form double helical stems (A helices) with a Kuhn length of about 140 nm [3, 4]. So, double stranded RNA is stiffer than double stranded DNA, which has a Kuhn length of 100 nm, noting that the Kuhn

length is twice the persistence length of a so-called wormlike chain.

The pairing of bases over long distances along the backbone gives rise to the secondary or folded structure of RNA. Pairing of bases can be represented by so-called arch diagrams. Nested arches represent helices, while crossings give rise to (the) so-called pseudoknots [5]. The nested pairings can be described quantitatively by recursion relations [6–8], which exactly sum all possible pairings without pseudoknots. From a geometrical point of view, the generated structures can be viewed as branched polymers. The size of an ideal, Gaussian linear polymer scales as the number of ‘segments’ to the power  $\nu = 1/2$ , while ideal branched ones have a scaling exponent  $\nu = 1/4$  [9]. Note that there is no excluded volume interaction between monomers of an ideal chain.

For self-avoiding chains the scaling exponents are  $\nu = 3/5$  and  $\nu = 1/2$  for the linear and branched polymers, respectively [9, 10]. However, because of its *tertiary* structures that include pseudoknots, RNAs are significantly more compact than branched polymers. Indeed, several numerical studies and surveys have found the exponent  $\nu = 1/3$  to be small for RNA, reflecting this more compact structure [11, 12].

Many small viruses encapsidate a single stranded RNA into a protein shell called the capsid. Under appropriate physico-chemical conditions of acidity and ionic strength, this process is spontaneous and the virus can readily assemble *in vitro* from solutions containing protein subunits and RNA [13–19]. Note that in the absence of genome, capsids do not form at physiological pH and salt concentrations. Many spherical viruses adopt structures with icosahedral symmetry [20, 21], which imposes a constraint on the number of subunits in capsids. The structural index  $T$ , introduced by Casper and Klug, defines the number of protein subunits in viral shells, which is 60 times the  $T$  number. Note that  $T = 1, 3, 4, 7, \dots$  can assume only certain ‘magic’ integer numbers [22–25].

Quite interestingly, virus protein subunits are able to co-assemble with a wide variety of negatively charged cargos, including non-cognate RNAs of different length and sequence, synthetic polyanions, and negatively charged nanoparticles [18, 26, 27]. It is now widely accepted that electrostatic interactions between the positive charges on the coat protein tails and negative charges on the cargo is the main driving force for the spontaneous assembly of simple viruses in solution [13–17, 28, 29]. Still, several recent self-assembly experimental studies reveal the importance of non-electrostatic interactions, associated with specific structures of the genome, for the selection of one RNA over another by the capsid proteins [30].

The self-assembly studies of Comas-Garcia *et al* [31] reveal in particular the importance of RNA topology. They carried out a number of experiments in which a solution of the capsid proteins of cowpea chlorotic mottle virus (CCMV) were mixed with equal amount of RNA1 of Brome mosaic virus (BMV) and RNA1 of Cowpea chlorotic mottle virus (CCMV). In this head-to-head competition, the amount of coat protein (CP) of CCMV was selected such that it could only encapsidate one of the genomes. Quite unexpectedly, the RNA1 of CCMV (the cognate RNA) lost to RNA1 of BMV, i.e. only RNA1 of BMV was encapsidated by CCMV CPs. These experiments emphasize the impact of RNA structure on the assembly of viral shells, as RNA1 of BMV has a more compact structure than that of CCMV [32].

Following these experiments a number of simulation studies, using quenched (fixed) branched polymers as a model for RNA, have shown that the optimal length of encapsidated RNA increases when accounting for its secondary structure [12, 33]. Mean-field calculations using annealed (equilibrium) branched polymers as model RNAs have also shown that the length of encapsidated polymer increases as the propensity to form larger numbers of branched points increases [32, 34, 35]. More importantly, these calculations show that a higher level of branching considerably increases the depth of the free-energy gain associated with the encapsulation of RNA by a positively charged shell. This implies that the efficiency



**Figure 1.** (a) The secondary structure of the CCMV RNA1 and (b) a random RNA with the same number of nucleotides. The structures are obtained using the Vienna RNA package [8].

of genome packaging goes up with increasing the level of branching, so with increasing compact secondary structure of the genome.

In fact, it was shown in [36, 37] that while RNA molecules of the same nucleotide length and composition might have similar amounts of base pairing, non-viral RNAs have significantly less compact structures than viral ones. The compactness of viral RNAs has been associated with the presence of a larger fraction of higher-order junctions or branch points in their secondary structure [36, 38, 39]. Figures 1(a) and (b) illustrate the secondary structures of CCMV RNA and those of a randomly sequenced RNA with the same length. The structures are obtained through the Vienna RNA software package [8]. As shown in the figure, CCMV RNA has considerably larger number of branched points than non-viral RNA of the same length.

Above-mentioned theoretical and experimental studies indicate that in a head-to-head competition between two different types of RNAs, the RNA with a larger number of branching junctions or branch points should have a competitive edge over others [32, 34, 35]. A naive physical explanation is that branching causes RNA molecules to become more compact than structureless linear polymers of similar chain length, which are then easier to accommodate in the limited space provided by the cavity of a capsid. According to these theories and simulations, a linear chain should definitely ‘lose’ to a branched one of the same number of monomers when competing head-to-head for a limited number of capsid proteins.

To probe the effect of RNA structure and test the above theories on the self-assembly of virions more systematically, Beren *et al* [40] recently performed a set of *in vitro* packaging experiments with polyU, an RNA molecule that has no folded

secondary structure. They examined whether RNA topology, i.e. the secondary structure or level of branching, allows the viral RNA to be exclusively packaged by its cognate capsid proteins. More specifically, they studied the competition between CCMV viral RNA with polyU of equal number of nucleotides for virus capsid proteins. They find that CCMV CPs are capable of packaging polyU RNAs and, quite interestingly, polyU outcompetes the native CCMV RNA in a head-to-head competition for the capsid proteins. These findings are in sharp contrast with the previous experimental, theoretical, simulation and scaling studies noted above, which suggest that the branching and compactness of RNA must lead to a more efficient capsid assembly. That being said, the scaling theory of [41] already hints at the subtle interplay between Kuhn length, solvent quality and linear charge density dictating the free energy gain of encapsulation.

To explain these intriguing experimental findings, we employ a mean-field density functional theory and study the impact of RNA branching, while allowing for differences in Kuhn length. We further consider that double helical sequences have a larger linear charge density than non-hybridized sequences along the chain. In all previous theoretical and simulation studies related to the impact of RNA topology on virus assembly, the focus has been on the importance of the degree of branching, ignoring the impact of base-pairing on the RNA Kuhn length and linear charge density.

As noted above, the Kuhn length of single stranded RNA under physiological conditions of monovalent salt is between one and two nm depending on the ionic strength [2], while that of a double stranded RNA is about 140 nm [3, 4]. The average duplex length of viral RNA is about six nucleotide pairs [11], which corresponds to about five nm. This value is much smaller than the persistence length of double stranded RNA [36], suggesting that viral RNA can be modeled as a flexible polymer with an average Kuhn length of about six paired nucleotides. There are of course also loop sequences that in our model act as end, hinge and branching points, but how this translates into an effective Kuhn length for the entire branched chain representation of the RNA is unclear. Plausibly, the effective Kuhn length of the internally hybridized chain should be larger than that of the equivalent unstructured non-hybridized chain. Furthermore, another major difference between the linear and branched (base-paired) ssRNA structures seems to be the linear charge density, which doubles for the latter on account of base pairing (hybridization).

In this paper, we vary the degree of branching as well as the effective Kuhn length and linear charge density of a model RNA, and study their impact on the optimal length of encapsulated genome by capsid proteins. We find that as we increase the chain stiffness or Kuhn length the free energy of encapsulation of RNA becomes less negative than that of a linear chain, at least under certain conditions. Hence, a larger Kuhn length, associated with base-pairing, might decrease the efficiency of packaging of RNA compared to a linear polymer. In contrast, our results indicate that increasing the linear charge density improves the efficiency of packaging of both linear and branched polymers. Thus base-pairing has two competing effects: it makes the chain stiffer, which increases the work

required to encapsidate the chain, but at the same time it increases the linear charge density that lowers the encapsidation free energy and augment the packaging efficiency. These results are consistent with the experiments of Beren *et al* [40], in which the linear RNA, PolyU, outcompetes the cognate RNA of CCMV when they are both in solution with a limited amount of capsid proteins of CCMV, that is, sufficient to encapsidate either PolyU or CCMV RNA but not both.

The remainder of this paper is organized as follows. In the next section, we introduce the model and present the equations that we will employ later. In section 3, we present our results and discuss the impact of the Kuhn length on the capsid stability and optimal length of encapsidated genome in section 4. Finally, in section 5, we present our conclusion and summarize our findings.

## 2. Model

To obtain the free energy associated with a genome trapped inside a spherical capsid, we consider RNA as a generic flexible branched polyelectrolyte that interacts with positive charges residing on the inner surface of the capsid. We focus on the case of annealed branched polymers as the degree of branching of RNAs, a statistical quantity, can be modified by its interaction with the positive charges on the capsid proteins [42].

Within mean-field theory, the free energy of a negatively charged chain in a salt solution confined inside a positively charged spherical shell can be written as [29, 32, 34, 35, 43–45]

$$\begin{aligned} \beta F = & \int d^3r \left[ \frac{\ell^2}{6} |\nabla \Psi(\mathbf{r})|^2 + \frac{1}{2} v \Psi^4(\mathbf{r}) + W[\Psi(\mathbf{r})] \right. \\ & \left. - \frac{1}{8\pi\lambda_B} |\nabla \beta e\Phi(\mathbf{r})|^2 - 2\mu \cosh [\beta e\Phi(\mathbf{r})] + \beta\tau\Phi(\mathbf{r})\Psi^2(\mathbf{r}) \right] \\ & + \int d^2r [\beta\sigma\Phi(\mathbf{r})] \end{aligned} \quad (1)$$

with  $\beta$  the inverse of temperature in units of energy,  $v$  the effective excluded volume per monomer,  $\lambda_B = e^2\beta/4\pi\epsilon$  the Bjerrum length,  $e$  the elementary charge,  $\mu$  the number density of monovalent salt ions, and  $\tau$  the charge of the statistical Kuhn segment of the chain. The dielectric permittivity of the medium  $\epsilon$  is assumed to be constant [46]. The quantity  $\ell$ , the Kuhn length of the polymer, is defined as an effective stiffness averaged over the entire sequence along the genome. Further, the fields  $\Psi(r)$  and  $\Phi(r)$  describe the square root of the monomer density field and the electrostatic potential, respectively, and the term  $W[\Psi]$  corresponds to the free energy density of an annealed branched polymer as described in equation (2) below.

As discussed in the Introduction, the secondary structure of the RNA molecules contain considerable numbers of junctions of single-stranded loops from which three or more duplexes exit. This makes RNA act effectively as a flexible branched polymer in solution. While the Kuhn length for a single stranded, non self-hybridized ssRNA is a few nanometers

and that for a double stranded RNA is about 140 nanometers, the Kuhn length of viral RNA is not well determined, as we discussed above. In the absence of exact measurements, we employ an average or effective value for  $\ell$ , which presumably will be larger if the number of consecutive base pairs (duplexes) between single stranded segments or stem loops along the RNA is larger. Further, we consider the limit of long chains consisting of a very large number of segments  $N \rightarrow \infty$  for our confined chains, where  $N$  denotes the number of segments. In this formal limit, we employ the ground-state dominance approximation implicit in equation (1), as it has proven to be accurate provided  $N \gg 1$ , i.e. for very long chains [47]. We specify below the connection between the number of segments and the number of nucleotides that make up the RNA, differentiating between self-hybridized and non self-hybridized RNAs.

The first term in equation (1) is the entropic cost of deviation from a uniform chain density and the second term describes the influence of excluded volume interactions. The last two lines of equation (1) are associated with the electrostatic interactions between the chain segments, the capsid and the salt ions at the level of Poisson-Boltzmann theory [43, 48, 49]. The term  $W[\Psi]$  represents the free energy density associated with the annealed branching of the polymer [50–53],

$$W[\Psi] = -\frac{1}{\sqrt{\ell^3}}(f_e\Psi + \frac{\ell^3}{6}f_b\Psi^3), \quad (2)$$

where  $f_e$  and  $f_b$  are the fugacities of the end and branched points of the annealed polymer, respectively [44]. Note that the stem-loop or hair-pin configurations of RNA are counted as end points. The quantity  $\frac{1}{\sqrt{\ell^3}}f_e\Psi$  indicates the density of end points and  $\frac{\ell^3}{6}f_b\Psi^3$  the density of branch points. The number of end  $N_e$  and branched points  $N_b$  are related to the fugacities  $f_e$  and  $f_b$ , respectively, and can be written as

$$N_e = -\beta f_e \frac{\partial F}{\partial F} \quad \text{and} \quad N_b = -\beta f_b \frac{\partial F}{\partial F}. \quad (3)$$

There are two additional constraints in the problem. Note first that the total number of monomers (Kuhn lengths) inside the capsid is fixed [54, 55],

$$N = \int d^3\mathbf{r} \Psi^2(\mathbf{r}). \quad (4)$$

We impose this constraint through a Lagrange multiplier,  $\mathcal{E}$ , introduced below. Second, there is a relation between the number of end and branched points,

$$N_e = N_b + 2, \quad (5)$$

as there is only a single polymer in the cavity that by construction has no closed loops as it has to mimic the secondary structure of an RNA. The polymer is linear if  $f_b = 0$ , and the number of branched points increases with increasing value of  $f_b$ . For our calculations, we vary  $f_b$  and find  $f_e$  through equations (3) and (5). Thus,  $f_e$  is not a free parameter.

Varying the free energy functional with respect to the monomer density field  $\Psi(r)$  and the electrostatic potential  $\Phi(r)$ , subject to the constraint that the total number of monomers

inside the capsid is constant [55], we obtain two self-consistent non-linear differential equations, which couple the monomer density with the electrostatic potential in the interior of the capsid. The resulting equations are

$$\frac{\ell^2}{6} \nabla^2 \Psi = -\mathcal{E}\Psi(\mathbf{r}) + \beta\tau\Phi(\mathbf{r})\Psi(\mathbf{r}) + \nu\Psi^3 + \frac{1}{2} \frac{\partial W}{\partial \Psi} \quad (6a)$$

$$\frac{\beta e^2}{4\pi\lambda_B} \nabla^2 \Phi_{in}(\mathbf{r}) = 2\mu e \sinh \beta e \Phi_{in}(\mathbf{r}) - \tau\Psi^2(\mathbf{r}) \quad (6b)$$

$$\frac{\beta e^2}{4\pi\lambda_B} \nabla^2 \Phi_{out}(\mathbf{r}) = 2\mu e \sinh \beta e \Phi_{out}(\mathbf{r}) \quad (6c)$$

with  $\mathcal{E}$  the earlier mentioned Lagrange multiplier enforcing the fixed number of monomers in the cavity. The boundary conditions for the electrostatic potential inside and outside of the spherical shell of radius  $R$  are

$$\hat{n} \cdot \nabla \Phi_{in} |_{r=R} - \hat{n} \cdot \nabla \Phi_{out} |_{r=R} = 4\pi\lambda_B\sigma/\beta e^2 \quad (7a)$$

$$\Phi_{in}(\mathbf{r}) |_{r=R} = \Phi_{out}(\mathbf{r}) |_{r=R} \quad (7b)$$

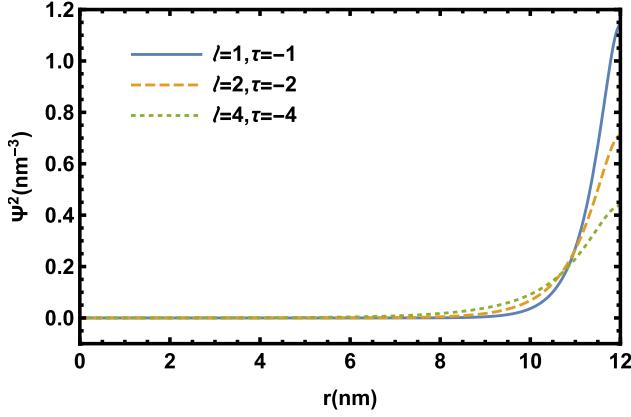
$$\Phi_{out}(\mathbf{r}) |_{r=\infty} = 0. \quad (7c)$$

The boundary condition (BC) for the electrostatic potential is obtained by minimizing the free energy assuming the surface charge density  $\sigma$  is fixed. The concentration of the polymer outside of the capsid is assumed to be zero. The BC for the inside monomer density field  $\Psi$  is of Neumann type ( $\hat{n} \cdot \nabla \Psi|_s = 0$ ) that can be obtained from the energy minimization [55] but our findings are robust and our conclusion do not change if we impose the Dirichlet boundary condition  $\Psi(r) |_{r=R} = 0$ . The former represent a neutral surface, whilst the latter a repelling surface [47].

### 3. Results

We solved the coupled equations given in equation (6) for the  $\Psi$  and  $\Phi$  fields, subject to the boundary conditions in equation (7) through a finite element method. The polymer density profiles  $\Psi^2$  as a function of the distance from the center of the shell,  $r$ , are shown in figure 2 for different values of the RNA stiffness  $\ell$  and a fixed number of nucleotides, presuming the RNA not to have any secondary structure. Note that for simplicity we assume that a linear chain with  $\ell = 1$  nm contains one nucleotide and carries one negative charge, so  $\tau = -e$ .  $\ell = 2$  nm has two nucleotides with two negative charges and so on. Thus in our figures the numerical value of  $\ell$  also indicates the number of nucleotides in one Kuhn length for linear chains. For the three plots in figure 2, the total number of nucleotides is calculated using equation (4) and is equal to 1000. It is worth mentioning that equation (4) gives us the total number of Kuhn lengths  $N$  and we multiply it by  $\ell$  the number of nucleotides along one Kuhn length to obtain the total number of nucleotides.

As illustrated in the figure, the polymer density becomes larger at the wall as the Kuhn length decreases, even though the linear charge density is fixed. In all plots for figure 2 we

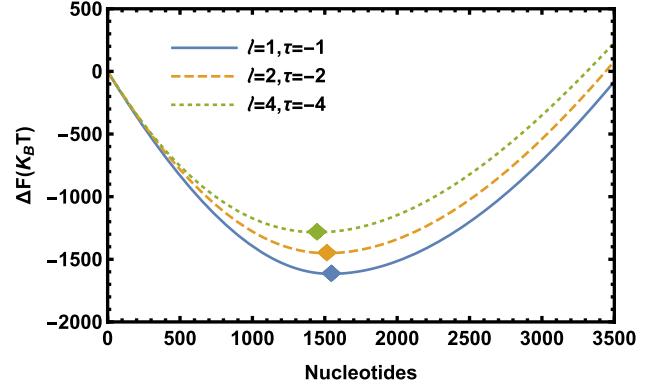


**Figure 2.** Genome density profile as a function of distance from the capsid center for a linear polymer with  $l = 1$  nm (solid line),  $l = 2$  nm (dashed line) and  $l = 4$  nm (dotted line). Other parameters used correspond to a  $T = 3$  virus: the total capsid charges on capsid  $Q_c = 1800e$ , the strength of excluded volume interaction  $v = 0.05 \text{ nm}^3$ , the fugacity  $f_b = 0$ , the quantity  $\mu$  corresponds to a salt concentration of 100 mM, the capsid radius  $R = 12 \text{ nm}$ , the temperature  $T = 300 \text{ K}$  and total number of nucleotides for all three cases equals 1000.

assumed that the excluded volume is kept constant. Arguably, the excluded volume parameter  $v$  depends on  $l$ , and usually it is assumed that  $v \propto l^3$  [47]. As we will discuss in section 4, our conclusions about the role of stiffness in the encapsidation free energy are robust and should not sensitively depend on the strength of the excluded volume interaction.

To investigate the packaging efficiency of a linear chain as a function of its stiffness, we obtained the free energy of the encapsidation of the linear polymer model as a function of number of nucleotides for different values of  $l$ , as illustrated in figure 3. The figure shows that the optimal number of nucleotides trapped in the shell increases as  $l$  decreases. We emphasize again that since we assumed that the size of a single nucleotide is about one nm, the numerical value of  $l$  represents the number of nucleotides within one Kuhn length. This implies that the number of nucleotides and hence the number of charges per Kuhn segment should increase as the Kuhn length increases. For example, in our parametrization  $l = 4$  nm represents four nucleotides (resulting in  $\tau = -4e$ ). We observe the same behavior for the free energy of branched polymers, that is, increasing  $l$  causes the optimal length of genome to move towards shorter chains. Obviously the stiffness value  $l$  is larger for the RNAs whose average number of base pairs in the duplex segments is larger.

The concept of the number of nucleotides per Kuhn length is trickier to implement for the branched polymers taken as model for self-hybridized ssRNA. For example, a branched polymer with the Kuhn length  $l = 1$  nm represents in our model description two nucleotides and a charge of  $\tau = -2e$ . When the average number of base pairs is about 8 in duplex segments of an ssRNA, we consider the Kuhn length is about 8 nm, but the number of nucleotides and number of charges per Kuhn length  $\tau$  will be 16. Thus, in our prescription of the self-hybridized ssRNA the number of nucleotides is twice the value of  $l$  within a Kuhn length as a result of base pairing.

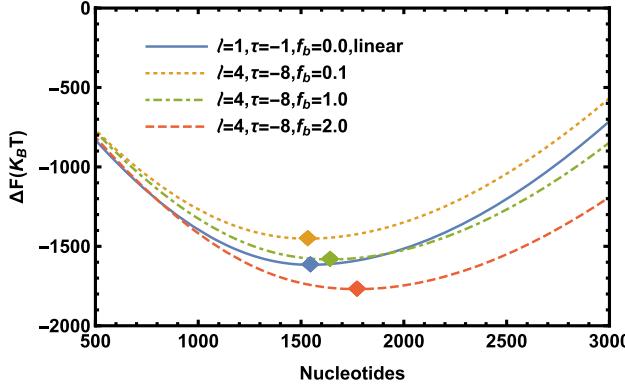


**Figure 3.** Encapsidation free energy of a linear polymer as a function of number of nucleotides for  $l = 1$  nm (solid line),  $l = 2$  nm (dashed line) and  $l = 4$  nm (dotted line). As the stiffness  $l$  increases, the optimal number of nucleotides moves towards shorter chains. The quantity  $\tau$  indicates the number of negative charges in one Kuhn segment. Other parameters used are the total number of charges on the capsid  $Q_c = 1800$ , the excluded volume parameter  $v = 0.05 \text{ nm}^3$ , the quantity  $\mu$  corresponds to a salt concentration of 100 mM, the radius of the cavity of the capsid  $R = 12 \text{ nm}$  and the absolute temperature  $T = 300 \text{ K}$ .

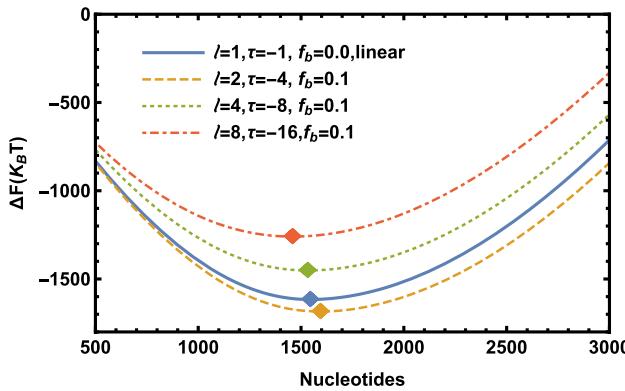
We also examined the impact of the fugacity on the optimal number of nucleotides. There is a direct relation between the fugacity and the number of branched points: As the fugacity increases the number of branched points of RNA increases too, see [32, 34, 35]. Figure 4 illustrates that the optimal number of nucleotides increases and the encapsidation free energy becomes more negative, indicating a more stable complex, as the fugacity of branching and hence the number of branch points increases. The solid line in the figure shows the free energy of a linear polymer. For the case shown in the figure, the Kuhn length of the linear chain is  $l = 1$  nm but that for the branched polymers  $l = 4$  nm, corresponding to four base-paired nucleotides. The number of charges within one Kuhn length then is  $\tau = -8e$ .

Figure 4 reveals that the free energy of the linear chain is lower than that of the branched one in certain regions of parameter space. For example, for a branched polymer with fugacity  $f_b = 0.1$ ,  $l = 4$  nm and  $\tau = -8e$  (dotted line), the encapsidation free energy of a linear chain with  $l = 1$  nm and  $\tau = -e$  is always lower than that of the branched polymer, and thus, in a head-to-head competition with a limited number of proteins, the linear chain will be the one that is preferentially encapsidated by capsid proteins. This shows that the work of compaction of linear chains could be lower than that of a branched polymer, depending on the stiffness and the degree of branching of the polymers involved. Note that for a fixed  $l$  while the number of branch points ( $f_b$ ) increases, at some point, the branched polymers outcompetes the linear polymer for binding to capsid proteins, as is illustrated in the figure.

We next studied the free energy of a branched polymer with a fixed fugacity for different values of the stiffness  $l$ . As illustrated in figure 5 for a fugacity  $f_b = 0.1$ , the linear chain (solid) ‘loses’ to a branched one when four nucleotides have formed two base pairs with  $l = 2$  nm and  $\tau = -4e$  (dashed line). However, the figure shows that as  $l$  increases, for  $l = 4$  nm



**Figure 4.** Encapsidation free energy as a function of number of nucleotides for a linear (solid line) and branched chains with different degree of branching:  $f_b = 0.1$  (dotted line),  $f_b = 1$  (dot-dashed line) and  $f_b = 2$  (dashed line). As the fugacity  $f_b$  (and hence the number of branched points) increases, the optimal number of nucleotides moves towards longer chains. Other parameters are  $Q_c = 1800e$ ,  $v = 0.05 \text{ nm}^3$ , the quantity  $\mu$  corresponds to a salt concentration of 100 mM,  $R = 12 \text{ nm}$  and  $T = 300 \text{ K}$ .



**Figure 5.** Encapsidation free energy as a function of number of nucleotides for a linear (solid line) and a branched chain at  $\ell = 2 \text{ nm}$  (dashed line),  $\ell = 4 \text{ nm}$  (dotted line) and  $\ell = 8 \text{ nm}$  (dot-dashed line). Other parameters used are  $Q_c = 1800e$ ,  $v = 0.05 \text{ nm}^3$ , the quantity  $\mu$  corresponds to a salt concentration of 100 mM,  $R = 12 \text{ nm}$  and  $T = 300 \text{ K}$ .

and 8 nm (dotted and dotted-dashed lines), their encapsidation free energies become larger than that of the linear chain, indicating that in a head-to-head competition the linear polymer will be encapsidated. Thus, if the average number of nucleotides in duplex segments increases, it becomes energetically more costly to confine RNA inside the capsid.

#### 4. Discussion

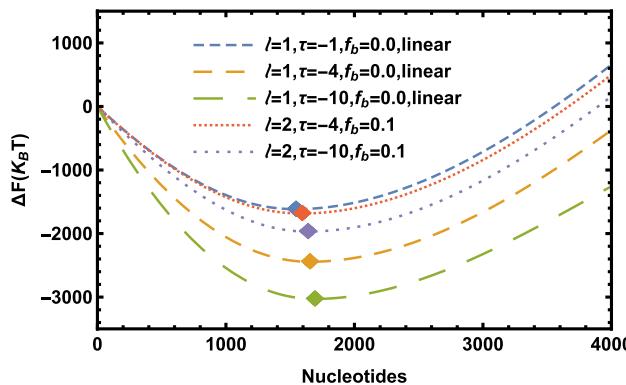
Recent experiments emphasized on the crucial role of the RNA topology in the efficiency of virus assembly. As noted in the introduction, Comas-Garcia *et al* [31] have shown that CCMV capsid proteins exclusively encapsidate BMV RNA in the presence of the cognate CCMV RNA under conditions where there is a limited number of capsid proteins in solution. The simulations and analytical studies performed in [32–35, 56] are consistent with these results: the viral RNA with a larger degree of branching has a competitive edge over the

other viral RNAs or non-viral randomly branched RNAs, keeping all other chain quantities equal.

Indeed, all mean-field theories, numerical calculations and simulations up to now have indicated that the encapsidation free energy of both annealed and quenched branched polymers is significantly lower than that of linear polymers. This suggests that if there are equal amounts of linear and branched polymers in a solution, but there are sufficient capsid proteins to encapsulate exclusively half of the genomes in solutions, only the branched polymer is encapsidated by capsid proteins. Nevertheless, according to a series of more recent experiments by Beren *et al* [40] in a head-to-head competition between a linear (polyU) chain and CCMV RNA of equal length, surprisingly, and in contrast to theoretical predictions, the linear chain outcompetes the cognate RNA.

While previous theoretical studies have focused on the scaling behavior of linear and branched flexible polymers [32, 34, 35, 48, 56–58], in this paper we study the impact of the stiffness or Kuhn length on the encapsidation of RNA by capsid proteins. In general the duplexed segments of viral RNA contain on average about five to six base-pairs [11]. Note that some studies show that viral RNAs must have between 60 and 70 % of their nucleotides in duplexes, so the linear charge density is almost a factor of two larger and the effective chain length about twice shorter [59]. We argue that while the base pairing on the one hand makes the RNA more compact, on the other hand it increases the effective Kuhn length or the statistical length of the polymer unit. This leads to an increase in the work of compaction of the flexible chain by capsid proteins, which is directly related to the encapsidation free energy of the polymer as plotted in figure 5. We emphasize again that the findings of this paper is not in contradiction with the previous studies: The more strongly branched a polymer is, the more competitive it becomes to be encapsidated by capsid proteins. However, in this work we show that because of base-pairing, the RNA also becomes stiffer and under appropriate conditions can no longer outcompete the linear polymer for binding to capsid proteins.

Since branching due to base-pairing causes both the stiffness and the linear charge density of an otherwise linear polymer to increase, one might wonder which effect, higher charge density or larger stiffness, makes the viral RNA less competitive than a linear polymer. Figure 6 distinguishes the effect of stiffness and charge density. The dashed lines in the figure correspond to linear polymers with  $\ell = 1 \text{ nm}$  but different numbers of charges per Kuhn segment  $\tau = -e, -4e, -10e$ . In the plots, the longer the dashes are, the higher the charge density is. As illustrated in the figure, the encapsidation free energy becomes lower as the charge density increases. The charge density has the same impact on the encapsidation free energy of branched polymers. Figure 6 shows that as the charge density of branched polymer increases (dotted lines), their free energy decreases. The more distance between the dots, the higher the charge density of the branched polymer. Quite interestingly, the figure shows that the effect of stiffness overshadows the impact of charge density. A branched polymer with the stiffness of  $\ell = 2 \text{ nm}$  and charge density of  $\tau = -4e$  or  $-10e$  has a higher free energy than a linear polymer with



**Figure 6.** Encapsidation free energy as a function of the number of nucleotides for linear (dashed lines) and branched chains (dotted lines), with various number of charges within one Kuhn length  $\tau$ . Other parameters are  $Q_c = 1800e$ ,  $v = 0.05 \text{ nm}^3$ , the quantity  $\mu$  corresponds to a salt concentration of 100 mM,  $R = 12 \text{ nm}$  and  $T = 300 \text{ K}$ .

the stiffness of  $\ell = 1 \text{ nm}$  but the charge density of  $\tau = -4e$ . These examples do not correspond to ‘real’ RNA as it is not possible to increase the number of charges to more than  $2e$  per base pair, but they clarify that base-pairing has three competing effects. First, it makes RNA stiffer, which increases the work of encapsidation but, second, in parallel gives rise to the branching effect and, third, a higher charge density, which both lowers the encapsidation free energy and enhances the packaging efficiency of RNA by capsid proteins.

Another important point to consider is the change in the excluded volume interaction that must somehow be connected with the variation in the Kuhn length. We repeated the calculations done for figure 5, but considered the excluded volume effect, which approximately goes as  $\ell^3$  [47]. We found that our conclusion is robust and that the excluded volume interaction only slightly modifies the boundary in the parameter space where the linear polymers are able to outcompete the branched ones. The results of this study can explain the intriguing findings of the experiments of Beren *et al* [40] in which the unstructured polyU RNA is preferentially packaged and outcompetes native RNA CCMV, despite the fact that viral RNAs have more branch points and as such have a more compact structure. Last but not least, we note that the interaction of RNA with capsid proteins could modify the preferred curvature of proteins and result into the capsid of different sizes and  $T$  numbers as demonstrated in [40]. However, since very little is known about this effect, in this paper we exclusively focused on the impact of RNA stiffness resulting from its base pairing in the RNA encapsidation free energy.

## 5. Conclusions

The results of our field theory calculations have shown that competition between different forms of RNA for encapsulation by virus coat proteins is a complex function of the degree of branching, effective stiffness of the polymer, linear charge density and excluded volume interactions. The conclusion of previous works that the more branched an RNA is on account of its secondary, base-paired structure, the

larger the competitive edge it has to be encapsulated in the presence of coat proteins needs to be refined. Under appropriate conditions of linear charge density and effective chain stiffness, we find that a linear chain may in fact outcompete even the native RNA of a virus, as was recently also shown experimentally. Of course, our conclusions are based on coarse-grained model in which the RNA binding domains of the coat proteins are represented by a smooth, uniformly charged wall. In future work we intend to more realistically model these polycationic tails that form a complex with the polynucleotide. Of particular interest here is the impact of excluded volume interactions between these tails and the polynucleotide.

## Acknowledgments

The authors would like to thank Jef Wagner for useful discussions. This work was supported by the National Science Foundation through Grant No. DMR-1719550 (RZ).

## ORCID iDs

Siyu Li  <https://orcid.org/0000-0002-4326-7560>

## References

- [1] Higgs P G 2000 *Q. Rev. Biophys.* **33** 199
- [2] Chen H, Meisburger S P, Pabit S A, Sutton J L, Webb W W and Pollack L 2012 *Proc. Natl Acad. Sci.* **109** 799
- [3] Kebbekus P, Draper D E and Hagerman P 1995 *Biochemistry* **34** 4354
- [4] Abels J A, Moreno-Herrero F, van der Heijden T, Dekker C and Dekker N H 2005 *Biophys. J.* **88** 2737
- [5] Nussinov R, Pieczenik G, Griggs J R and Kleitman D J 1978 *SIAM J. Appl. Math.* **35** 68
- [6] McCaskill J S 1990 *Biopolymers* **29** 1105
- [7] Zuker M 2003 *Nucleic Acids Res.* **31** 3406
- [8] Hofacker I L, Fontana W, Stadler P F, Bonhoeffer L S, Tacker M and Schuster P 1994 *Mon. hefte Chem.* **125** 167
- [9] Schwab D and Bruinsma R F 2009 *J. Phys. Chem. B* **113** 3880
- [10] Grosberg A Y 1997 *Phys.—Usp.* **40** 125
- [11] Fang L T, Gelbart W M and Ben-Shaul A 2011 *J. Chem. Phys.* **135** 155105
- [12] Singaram S W, Garmann R F, Knobler C M, Gelbart W M and Ben-Shaul A 2015 *J. Phys. Chem. B* **119** 13991
- [13] Sikkema F D, Comellas-Aragones M, Fokkink R G, Verduin B J M, Cornelissen J and Nolte R J M 2007 *Org. Biomol. Chem.* **5** 54
- [14] Ren Y P, Wong S M and Lim L Y 2006 *J. Gen. Virol.* **87** 2749
- [15] Ni P, Wang Z, Ma X, Das N C, Sokol P, Chiu W, Dragnea B, Hagan M and Kao C C 2012 *J. Mol. Biol.* **419** 284
- [16] Losdorfer Bozic A, Siber A and Podgornik R 2013 *J. Biol. Phys.* **39** 215
- [17] Zlotnick A, Aldrich R, Johnson J M, Ceres P and Young M J 2000 *Virology* **277** 450
- [18] Sun J *et al* 2007 *Proc. Natl Acad. Sci. USA* **104** 1354
- [19] Ning J, Erdemci-Tandogan G, Yufenyuy E L, Wagner J, Himes B A, Zhao G, Aiken C, Zandi R and Zhang P 2016 *Nat. Commun.* **7** 13689
- [20] Fejer S, Chakrabarti D and Wales D 2010 *ACS Nano.* **4** 219
- [21] Rapaport D C 2004 *Phys. Rev. E* **70** 051905
- [22] Wagner J and Zandi R 2015 *Biophys. J.* **109** 956

[23] Luque A, Zandi R and Reguera D 2010 *Proc. Natl Acad. Sci.* **107** 5323

[24] Chen T, Zhang Z and Glotzer S C 2007 *Proc. Natl Acad. Sci.* **104** 717

[25] Paquay S, Kusumaatmaja H, Wales D J, Zandi R and van der Schoot P 2016 *Soft Matter* **12** 5708

[26] Kusters R, Lin H K, Zandi R, Tsvetkova I, Dragnea B and van der Schoot P 2015 *J. Phys. Chem. B* **119** 1869

[27] Hagan M F and Zandi R 2016 *Curr. Opin. Virol.* **18** 36

[28] Lin H K, van der Schoot P and Zandi R 2012 *Phys. Biol.* **9** 066004

[29] Sivanandam V, Mathews D, Garmann R, Erdemci-Tandogan G, Zandi R and Rao A L N 2016 *Sci. Rep.* **6** 26328

[30] Patel N, Dykeman E C, Coutts R H A, Lomonosoff G P, Rowlands D J, Phillips S E V, Ranson N, Twarock R, Tuma R and Stockley P G 2015 *Proc. Natl Acad. Sci.* **112** 2227

[31] Comas-Garcia M, Cadena-Nava R D, Rao A L N, Knobler C M and Gelbart W M 2012 *J. Virol.* **86** 12271

[32] Erdemci-Tandogan G, Wagner J, van der Schoot P, Podgornik R and Zandi R 2014 *Phys. Rev. E* **89** 032707

[33] Perlmutter J D, Qiao C and Hagan M F 2013 *eLife* **2** e00632

[34] Erdemci-Tandogan G, Wagner J, van der Schoot P, Podgornik R and Zandi R 2016 *Phys. Rev. E* **94** 022408

[35] Li S, Erdemci-Tandogan G, Wagner J, Van Der Schoot P and Zandi R 2017 *Phys. Rev. E* **96** 022401

[36] Yoffe A M, Prinsen P, Gopal A, Knobler C M, Gelbart W M and Ben-Shaul A 2008 *PNAS* **105** 16153

[37] Bruinsma R F, Comas-Garcia M, Garmann R F and Grosberg A Y 2016 *Phys. Rev. E* **93** 1

[38] Li Tai F, Gelbart W M and Ben-Shaul A 2011 *J. Chem. Phys.* **135** 155105

[39] Gopal A, Egecioglu D E, Yoffe A M, Ben-Shaul A, Rao A L N, Knobler C M and Gelbart W M 2014 *PLoS ONE* **9** e105875

[40] Beren C, Dreesens L L, Liu K N, Knobler C M and Gelbart W M 2017 *Biophys. J.* **113** 339

[41] van der Schoot P and Zandi R 2013 *J. Biol. Phys.* **39** 289

[42] McPherson A 2005 *BioEssays* **27** 447

[43] Borukhov I, Andelman D and Orland H 1998 *Euro. Phys. J. B* **5** 869

[44] Wagner J, Erdemci-Tandogan G and Zandi R 2015 *J. Phys.: Condens. Matter* **27** 495101

[45] Erdemci-Tandogan G, Wagner J, van der Schoot P and Zandi R 2016 *J. Phys. Chem. B* **120** 6298

[46] Janssen M, Härtel A and van Roij R 2014 *Phys. Rev. Lett.* **113** 268501

[47] de Gennes P G 1979 Cornell University Press

[48] Siber A and Podgornik R 2008 *Phys. Rev. E* **78** 051915

[49] Shafir A, Andelman D and Netz R R 2003 *J. Chem. Phys.* **119** 2355

[50] Lubensky T C and Isaacson J 1979 *Phys. Rev. A* **20** 2130

[51] Nguyen T T and Bruinsma R F 2006 *Phys. Rev. Lett.* **97** 108102

[52] Lee S I and Nguyen T T 2008 *Phys. Rev. Lett.* **100** 198102

[53] Elleuch K, Lequeux F and Pfeuty P 1995 *J. Phys. I France* **5** 465

[54] de Gennes P G 1982 *Macromolecules* **15** 492

[55] Ji H and Hone D 1988 *Macromolecules* **21** 2600

[56] Erdemci-Tandogan G, Orland H and Zandi R 2017 *Phys. Rev. Lett.* **119** 188102

[57] Zandi R and van der Schoot P 2009 *Biophys. J.* **96** 9

[58] Siber A, Zandi R and Podgornik R 2010 *Phys. Rev. E* **81** 051919

[59] Borodavka A, Singaram S, Stockley P, Gelbart W, Ben-Shaul A and Tuma R 2016 *Biophys. J.* **111** 2077