Impact of the topology of viral RNAs on their encapsulation by virus coat proteins

Paul van der Schoot · Roya Zandi

Received: 23 January 2013 / Accepted: 7 February 2013 / Published online: 14 March 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Single-stranded RNAs of simple viruses seem to be topologically more compact than other types of single-stranded RNA. It has been suggested that this has an evolutionary purpose: more compact structures are more easily encapsulated in the limited space that the cavity of the virus capsid offers. We employ a simple Flory theory to calculate the optimal amount of polymers confined in a viral shell. We find that the free energy gain or more specifically the efficiency of RNA encapsidation increases substantially with topological compactness. We also find that the optimal length of RNA encapsidated in a capsid increases with the degree of branching of the genome even though this effect is very weak. Further, we show that if the structure of the branching of the polymer is allowed to anneal, the optimal loading increases substantially.

Keywords Virus assembly · Self-assembly · Branched polymers · Flory theory

1 Introduction

A simple spherical virus consists of a protein shell known as the capsid that contains the viral genome, typically a single-stranded (ss) RNA molecule [1]. The capsid itself is a supramolecular assembly of multiple copies of the same or a few structurally similar coat proteins. Because of the underlying icosahedral symmetry of the arrangement of the proteins in the shell, the number of coat proteins making up virus shells equals 60 times the so-called

P. van der Schoot
Group Theory of Polymers and Soft Matter, Eindhoven University of Technology,
PO Box 513, 5600 MB, Eindhoven, The Netherlands

P. van der Schoot
Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE, Utrecht, The Netherlands

R. Zandi (✉)
Department of Physics & Astronomy, University of California, Riverside, CA 92521, USA
e-mail: royaz@ucr.edu

T number of the virus, with $T = 1, 3, 4, 7, \dots$ and so on. Every virus has a specific T number associated with its well-defined size, although a few viruses come in more than a single T number [2].

A fair number of simple viruses have been reconstituted *in vitro* [2–11]. This includes rod-like viruses such as tobacco mosaic virus TMV [3, 4], the first virus reconstituted *in vitro*, and spherical viruses such as cowpea chlorotic mottle virus CCMV [8], the first spherical virus reassembled *in vitro* from its constituents. In fact, under the right conditions, virus coat proteins quite readily encapsulate the RNA of other viruses and even synthetic polyanions and charged nanoparticles [7–9, 12–15]. Interestingly, the purified capsid proteins of human immunodeficiency virus (HIV) also spontaneously assemble around the RNA of the virus, but in this case the formed capsids have a conical structure [16–19, 21–24].

The spontaneous reconstitution of viruses has led to the idea that virus assembly is a thermodynamic process [25], driven primarily (but not solely [26]) by electrostatic interactions between positive charges on RNA binding domains on the coat proteins and negative charges on the RNA molecules [27, 28]. This has spurred a flurry of theoretical work seeking to address questions relating to the distribution of RNA inside capsids and to fact that the total number of positive charges on the RNA binding domains, Q , that is linearly correlated with the number of charges on their ssRNA, M [29–39]. The latter equals the number of nucleotides of the genome. Recently, Belyi and Muthukumar [30] found that $M \simeq 1.6 \times Q$, implying that more charges are bound than necessary for charge neutralization. In polymer physics, this phenomenon is known as overcharging and is attributed to correlation effects between ions interacting with a charged surface, in this case caused by the connectivity of the charged polymeric chain [40].

For polyions such as RNA, overcharging is intuitive because charges that are away more than a Debye screening length or so from the charges on the RNA binding domains do not interact with them. These disordered domains, also known as arginine-rich motifs or ARMs, penetrate a few nanometers into the spherical cavity [30, 32]. The phenomenon of overcharging remains hotly debated in the polymer physics community and, not surprisingly, has lead to the application of quite different modeling approaches in the context of virus assembly [29–38, 41]. Intriguingly, virtually all of the calculations and computer simulations published to date rely on *linear* polyelectrolyte models representing the polynucleotide. An exception is the work of Lee and Nguyen, whose focus was on the radial density of RNA inside the capsid [42].

Even though single-stranded RNAs are linear polymers, they in fact resemble branched polymers in solution due to base pairing between complementary nucleotides that may or may not be far from each other along the primary sequence of the chain. In a groundbreaking study, de Gennes showed that the radius of gyration, R , of a model poly(dTA) RNA scales with molecular weight, M , as $R \sim M^{1/d_f}$ with a fractal dimension of $d_f = d_G = 4$ [43]. The Gaussian fractal dimension d_G is that of an ideal fractal object, i.e., one where excluded-volume effects are ignored. Since the fractal dimension of randomly branched polymers is also four [44–46], they are often thought to represent the statistical properties of ssRNAs in solution. Note that ideal randomly branched polymers are much more compact than ideal linear chains for the same molecular weight, because for the former $d_G = 4$ while for the latter $d_G = 2$.

More recent theoretical work on modeling ssRNAs give Gaussian fractal dimensions $d_G = 3 - 4$, depending on the temperature and nucleotide sequence [47–50]. Random nucleotide sequences, for instance, give $d_G = 3$ [50]. Base-pairing studies suggest that viral

RNA sequences are unusual and might even have a Gaussian fractal dimension as large as five [49]. The compactness of viral RNAs has been linked to the presence of a large fraction of higher-order junctions in their secondary structure, i.e., junctions of functionality larger than three (Gopal et al., submitted). The primary sequence of viral RNAs may well have evolved to encode not only for proteins but also to attain as compact structures as possible, hence resulting to the secondary and tertiary structures. This would plausibly facilitate encapsulation by coat proteins and provide a competitive edge in their competition with cellular RNAs for encapsulation in infected cells [49].

The connection between Gaussian fractal dimension and the actual fractal dimension and the physical size of RNA is actually not all that straightforward. In fact, it is not even clear if single-stranded RNAs of viruses are indeed fractals, and if they are, whether they are indeed much more compact than other types of RNA with larger Gaussian fractal dimension. A simple estimate based on the application of Flory theory for *self-avoiding* fractal objects predicts a simple relation between the actual fractal and Gaussian fractal dimensions, $d_f = 5d_G/2(d_G + 1)$ [51], at least in three spatial dimensions. This then gives a fractal dimension d_f of $5/3 \simeq 1.67$ for a linear chain with $d_G = 2$, and of $15/8 \simeq 1.88$, 2 and $25/12 \simeq 2.08$ for branched ones presuming $d_G = 3, 4$ and 5 , respectively. The larger d_G the more compact the object, as expected, but excluded-volume interactions strongly diminish the impact of branching on the size of the object. In fact, the impact of branching on the physical size of a polymer becomes smaller still, if it is allowed to restructure or *anneal* in response to volume exclusion. For a randomly quenched branched chain that in essence consists only of branch points and end points and for which $d_G = 4$, we have $d_f = 13/7 \simeq 1.86$, at least within Flory theory [45, 52]. Single-stranded RNAs are thought to attain a different secondary structure inside of a capsid than in free solution [4], so they should behave in a way closer to an annealed branched polymer than a quenched one, i.e., one with a fixed branching.

As already noted, the focus of the work of Lee and Nguyen [42] was on the concentration profile of an *annealed* branched polymer in the spherical cavity of a model capsid, where the interaction with ARM region was modeled by a delta-peaked attractive interaction with the wall. No optimal loadings or free energies were presented as a function of the level of branching of RNA. Here, we study a much simpler and mathematically more tractable model based on a Flory theory for the encapsulation of fractal objects with arbitrary Gaussian fractal dimension. We extend our earlier work [36], focused on quenched polymers with Gaussian dimensions $d_G = 2$ and $d_G = 4$, to calculate the free energy and optimal chain encapsulate for RNAs with arbitrary Gaussian dimensions. Further, we now explicitly consider the effects of annealed branching for the case of $d_G = 4$.

We find that the optimal length of encapsulated polymer increases only weakly with Gaussian fractal dimension for given attraction to the ARM region of the capsid that we model by way of a square well potential. This means that branching does not significantly increase the optimal loading of the capsid. The equilibrium thickness of the adsorbed fractal polymer does decrease with d_G , in agreement with experimental observations on encapsulation of linear polymers and RNAs [4, 12]. The most significant result that we obtain is that the free energy gain upon encapsulation increases significantly with increasing fractal dimension, by as much as 60% on going from $d_G = 3$ (for random RNA sequences) to $d_G = 5$ (the upper limit for viral RNAs) [48, 49] (Gopal et al., submitted). For $d_G = 4$, annealing may increase the absorbed amount by as much as 30%. Our studies indicate that it is the structure rather than the length of RNA that has a huge impact on the free energy and in consequence on the efficiency of RNA packaging by capsid proteins, which is consistent with the recent studies of Stockley and coworkers on virus assemblies [26].

The remainder of this paper is organized as follows. In Section 2, we present our three-zone adsorption model and associated mean-field free energy. For this free energy we consider the fractal object (“manifold”) in free solution as a reference state. We optimize the free energy in terms of the thickness of the adsorbed layer and the total length of the polymer and find that the references free energy only weakly influences the outcome of our calculations. For weak interactions we are able to present analytical results. In Section 3 we discuss our results and present in Section 4 our main conclusions.

2 Theory

In our model, the RNAs are represented by polymeric objects of Gaussian fractal dimension d_G . We presume the branching to be quenched, i.e., fixed, and postpone a discussion of how to incorporate annealing into the model to the end of this section. The polymer segments of size l are thought to interact through a second virial coefficient (“excluded volume”) $\nu > 0$ that includes the effects of electrostatic repulsion between them. For single-stranded RNAs, we expect $l \simeq 1 - 2$ nm [29, 49, 50]. Let the number of segments of the polymer be equal to $M \gg 1$. We need to consider the free energy difference $\Delta F \equiv F_{\text{cap}} - F_{\text{free}}$ between the polymer confined in a capsid and that in free solution, described by the free energies F_{cap} and F_{free} , respectively. Electron microscopic images show that inside the capsid the genome forms a dense layer and seems to be adsorbed to the inner surface of the protein shell, where it makes contact with the positively charged ARMs of the coat proteins [30, 37].

For entropy reasons, we expect the layer thickness of the adsorbed polymer, D , to be larger than the thickness of the region, b , where the ARMS are located in the interior of the capsid [28, 38]. For the plant virus CCMV, we have the inner radius of capsid equal to $R_c \simeq 10$ nm [27]. There is no precise value for the size of ARM region in the literature but, fortunately, b drops out the final resulting equations and as such we do not need to know its exact value.¹ The polymer layer thickness $D \leq R_c$ depends on the strength of the interaction between RNA segments and the RNA binding domains, described by a phenomenological free energy parameter ε (in units of thermal energy), the overall length of the polymer M and the degree of branching as expressed in the Gaussian fractal dimension d_G . The free energy gain $\varepsilon > 0$ of a polymer segment in the ARM region depends on the balance between attractive and repulsive interactions with the ARM. This includes the effect of volume exclusion, electrostatic attraction, and so on. Below we specify it for a particular model but for now we keep it as a free parameter driving the assembly of the virus particles.

In the spirit of Flory theory, we write down the free energy of the polymer inside the capsid as

$$F_{\text{cap}} = c_1 \left(\frac{M^{1/d_G} l}{D} \right)^{d_G} - M \ln \left(1 - c_2 \frac{M \nu}{R_c^2 D} \right) - c_3 \varepsilon \left(\frac{b M}{D} \right), \quad (1)$$

in units of thermal energy, where c_i for $i = 1, 2, 3$ are numerical constants. This expression applies to three spatial dimensions. The first term describes the configurational entropy loss due to compression of the Gaussian reference polymer in a spherical annulus of width $D \leq R_c$ with radius R_c . It is accurate provided $M^{1/d_G} l \gg D$, and we verify this condition

¹The thickness of the ARM region depends on whether it makes contact with polyanionic cargo. See, e.g., [32].

a posteriori. The expression can be obtained from a simple scaling Ansatz and formally derived within the so-called ground-state approximation for $d_G = 2$ and $d_G = 4$ from Edwards-Lifshitz theory for linear chains [53] and extension of it to randomly branched chains [44]. In principle, c_1 could depend on d_G but we for simplicity choose to ignore this. The other constants should be independent of d_G , because they refer to interactions between pairs of individual segments.

The second term describes the volume exclusion between the segments, including the influence of screened electrostatic repulsion between them. It should be noted (i) that $M/R_c^2 D$ is approximately the number density of RNA segments in the polymer layer, (ii) that $\nu M/R_c^2 D$ must be proportional to a volume fraction in that layer, and (iii) that for low volume fraction we retrieve the familiar second-virial free energy (in units of thermal energy), $M\nu M/R_c^2 D$. Hence, the second term in (1) goes beyond the usual second-virial treatment typical of Flory theory, but for simplicity we do approximate the volume fraction by replacing the exact volume of the spherical annulus, proportional to $R_c^3 - (R_c - D)^3$, by $R_c^2 D$. For our purposes this suffices, as this is accurate in both interesting limits of $D \ll R_c$ and $D = R_c$.

The last term of (1) accounts for the fraction of Mb/D of the segments interacting with the ARMs. As discussed, the ARMS are disordered domains located mostly on the C-terminal end but sometimes on the N-terminal end of the coat proteins [1]. These can be viewed as flexible linear polymers even though upon complexation with RNA, a short section may become α helical such as is the case with CCMV and BMV [54]. If we treat the ARM as a linear polymer and presume that the collection of ARMs lining the inner surface of the shell behave like a polymer brush [30], we find again from Flory theory that the interaction free energy per RNA segment must be equal to $\varepsilon = -\nu_{\pm} n/ab$. Here, n is the number of segments making up an ARM, $\nu_{\pm} < 0$ is the second-virial coefficient between a RNA and an ARM segment and a the surface area per protein in a capsid. The latter is dominated by the attractive screened Coulomb interaction between them.

The free energy of the branched chain in free solution, F_{free} , is equivalent to that of (1), except that

$$F_{\text{free}} = c_4 \left(\frac{R}{M^{1/d_G} l} \right)^2 - M \ln \left(1 - c_5 \frac{M\nu}{R^3} \right), \quad (2)$$

where c_4 and c_5 are again constants of order unity. Note that here we plausibly presume that $R \gg M^{1/d_G} l$, i.e., excluded volume interactions swell the chain.

The optimal values of D and R minimize $\Delta F = F_{\text{cap}} - F_{\text{free}}$. As we are interested in the optimal amount of RNA encapsulated, we also minimize the free energy with respect to M . The optimal values of the adsorption layer thickness D_* , size of the free chain R_* and molecular weight M_* we insert back into ΔF , in order to find the free energy of encapsulation ΔF_* . Associated with it is a Boltzmann factor that turns out to describe a critical mole fraction of coat proteins below which the ssRNAs and protein remain free in solution and above which co-assembly takes place, at least if the assembly is sufficiently co-operative [27]. The assembly is co-operative if the coat proteins interact strongly enough in the complete shell: this produces a significant rim tension in an incomplete shell that suppresses the intermediate structures. In that case, we only have appreciable concentrations of naked RNAs and/or completely encapsulated ones [36].

We can derive a similar free energy to (1) for an *annealed* randomly branched polymer with $d_G = 4$. This is most simply explained for a chain in free solution. Due to annealing of the branching, there is an additional entropy contribution proportional to L^2/M relating

to the number of segments L separating two endpoints, and the entropy loss upon stretching the (Gaussian) chain connecting those endpoints that scales as $R^2/l^2 L$ [46, 52]. Because L couples directly to the configurational entropy and only indirectly to the interactions between the segments, L optimizes the total free energy if $L \approx M^{1/3} R^{2/3} l^{-2/3}$. This gives a configurational free energy that scales as $M(R/Ml)^{4/3}$. Hence, we obtain for the free energy of our self-avoiding annealed randomly branched chain in free solution,

$$F_{\text{free}} = c_6 M \left(\frac{R}{Ml} \right)^{4/3} - M \ln \left(1 - c_7 \frac{Mv}{R^3} \right), \quad (3)$$

as before in units of thermal energy with c_6 and c_7 constants. Here, we presumed that $R \gg M^{1/d_G} l$ with $d_G = 4$.

Applying similar scaling arguments to a chain in a spherical annulus, we find

$$F_{\text{cap}} = c_8 M \left(\frac{R_c^2}{MDl} \right)^{4/3} - M \ln \left(1 - c_9 \frac{Mv}{R_c^2 D} \right) - c_{10} \varepsilon \left(\frac{bM}{D} \right), \quad (4)$$

with $M^{1/d_G} l \gg D$, and c_8 and c_9 constants. Comparing (1)–(4), it seems reasonable to presume that $c_7 = c_5$, $c_9 = c_2$ and $c_{10} = c_3$. Note that the reshuffling of the branching is different in the chain in free solution from that in the capsid. The reason is that in one case the Gaussian chain is expanded while in the other it is confined (compressed). This leads to different fractal dimensions in the two cases.

The equilibrium values of D_* , M_* and R_* follow from the minimization of $\Delta F = F_{\text{cap}} - F_{\text{free}}$. Inserting the equilibrium values of these quantities back into ΔF , gives the free-energy gain upon assembly, ΔF_* . Analytical results for all relevant quantities are presented in the next section for varying Gaussian fractal dimension. We also directly compare results for $d_G = 4$ for the cases of quenched and annealed branching.

3 Results and discussion

The optimal adsorption layer thickness D_* and size R_* in free solution for the optimal molecular weight M_* of encapsulated polymer that we find for the quenched case obey highly non-linear coupled equations that we do not reproduce here. It turns out, however, that in the limit of large enough capsids $R_c \gg l$ and for sufficiently weak interactions $\varepsilon b/l < 1$ we are able to obtain analytical results. For instance, in this limit we obtain for the optimal polymer size M_* encapsulated by the coat proteins,

$$M_* \approx \frac{d_G - 1}{2d_G - 1} \times \left(\frac{\varepsilon b R_c^2}{v} \right), \quad (5)$$

where the constant of proportionality c_3/c_2 is omitted. Equation (5) shows that the larger the interaction, the longer the optimal length of the encapsulated polymer. Also, longer chains are encapsulated as the level of branching increases at fixed interaction strength although the effect is quite weak. Indeed, $M_*(d_G = 5)/M_*(d_G = 3) = 10/9 \approx 1.1$, so the length of viral RNAs packaged is at most a mere 10% more than that of RNAs with a random nucleotide sequences. This implies that nucleotide sequence optimization as discussed in the introduction does not significantly help to increase the amount of genome encapsulation. For large values of $\varepsilon b/l \gg 1$, the conformational entropy becomes subdominant and the physics of the problem is dictated by the balance between attractive and repulsive interactions. In that regime, the fractal dimension of the chain becomes irrelevant.

Another important (observable) quantity is the optimal adsorption layer thickness for the encapsulated polymer with the optimal length M_* ,

$$D_* \approx l \left((2d_G - 1) \times \frac{\varepsilon b}{l} \right)^{1/(d_G-1)}, \quad (6)$$

where we have dropped a constant of proportionality $(c_1/c_3)^{1/(d_G-1)}$. In order for (6) to take on reasonable values $b \leq D_* \leq R_c$, we need to require that $l/b(l/R_c)^{d_G-1} \lesssim \varepsilon \lesssim (l/b)^{d_G}$. This puts fairly tight bounds on the strict applicability of the theory, in particular for small viruses. Granting these limits, D_* increases with increasing interaction strength, ε , between the segments of the polyanionic cargo and the polycationic ARM of the coat protein. This is due to the fact that as the strength of interaction increases, the optimal chain size increases and, because of this, the optimal thickness, D_* , increases. This has to be contrasted with what we find when the chain length M is fixed: in that case the layer thickness D decreases as ε increases, as expected. D_* also decreases with increasing d_G , i.e., with increasing degree of branching. So, the more strongly branched a polymer is, the more localized the adsorbed polymer is in the capsid. This is to be expected and agrees with what is seen in electron microscopic images of virus capsids containing linear polyanions and those containing ssRNA.

If we insert our estimate $\varepsilon b/l \approx -v_{\pm} n/al$ into (6), we find that it depends in essence only on the number of segments, n , and surface density $1/a$ of the individual ARM, the fractal dimension of the RNA and the strength of the interaction between the charges on the RNA and the ARMs as described by the parameter $v_{\pm} < 0$. At the level of a Debye-Hückel approximation, we expect $v_{\pm} = v_{\pm}^0 - 4\pi\alpha_{+}\alpha_{-}\lambda_B\lambda_D$, where $v_{\pm}^0$ is the “bare” excluded volume between the structural units of the ssRNA and the ARMs, and α_{+} and α_{-} are the number of charges per segment on the ARM (+) and ssRNA (−). Finally, $\lambda_B \approx 0.7$ nm denotes the Bjerrum length and $\lambda_D \approx 0.3/\sqrt{c_S}$ nm the Debye length of the aqueous solution, with c_S the molar concentration of 1-1 salt [51]. The larger the salt concentration, the smaller ε and D_* .

Another conclusion that we can draw from our analysis is that

$$M_* \approx \left(\frac{d_G - 1}{2d_G - 1} \right) \times \left(\frac{-v_{\pm}}{\alpha_{+}v} \right) \times Q, \quad (7)$$

if we insert our estimate for ε taken from the polymer brush model for the ARM region in (5) and again ignore the constant of proportionality. Here, Q is the total number of charges on the ARMs in the cavity of the capsid. So, we do find that $M \propto Q$ as discussed in the introductory section, except that the constant of proportionality seems to depend both on the level of branching of the polyanion *and* on the solution conditions via the excluded volumes $v_{\pm} < 0$ and $v > 0$. This means that more polymer can be encapsulated by reducing the solvent quality and hence v , e.g., by adding *multivalent* cations that tend to condense ssRNAs by effectively lowering v [55]. Adding simple 1-1 salt alters $v_{\pm}$ and v in the same way, that is, both should become smaller in magnitude with increasing salt.

Indeed, for sufficiently low concentrations of added salt, we expect $v_{\pm} = v_{\pm}^0 - 4\pi\alpha_{+}\alpha_{-}\lambda_B\lambda_D \approx -4\pi\alpha_{+}\alpha_{-}\lambda_B\lambda_D$ to approximately hold, while $v = v^0 + 4\pi\alpha^2\lambda_B\lambda_D \approx 4\pi\alpha^2\lambda_B\lambda_D$. Here, $v_{\pm}^0$ and v^0 denote the “bare” volume exclusions between the various species. However, we do expect $v_{\pm}$ to change sign at sufficiently large ionic strength as electrostatic attraction becomes subdominant relative to, say, the effects of steric repulsion. This implies that at sufficiently large ionic strength, RNAs would not be encapsulated, which agrees with experimental observation [6]. All of this begs the question of whether the

seemingly constant overcharging of 60% that was found by Belyi and Muthukumar [30] for a large number of viruses is real or just a coincidence. From our admittedly crude analysis, it certainly seems a coincidence, not least because the ionic conditions in different types of cell are not likely to be exactly the same. Our theory also explains why in encapsulation experiments of synthetic linear polymers the overcharging ranges from 0.6 and to 9 [12–14].

We have seen that branching does not significantly increase the optimal amount of encapsulated RNA. What branching does do, however, is increase the efficiency of encapsulation. In the limit $R_c \gg l$ the free energy of encapsulation $\Delta F \approx F_{\text{cap}}$ is dominated by the free energy of the polymer in the capsid, giving

$$\Delta F \approx -\frac{(d_G - 1)^2}{(2d_G - 1)^{(2d_G - 1)/(d_G - 1)}} \times \left(\frac{R_c^2 l}{v}\right) \times \left(\frac{\epsilon b}{l}\right)^{(2d_G - 1)/(d_G - 1)} \quad (8)$$

apart from a constant of proportionality $(c_1^2/c_2) \times (c_3/c_1)^{(2d_G - 1)/(d_G - 1)}$ that we have dropped. According to (8), the free-energy gain of encapsulation $|\Delta F|$ is an increasing function of the strength of the adsorption energy ϵ , but less so as the degree of branching d_G increases. Still, if we take the earlier discussed cases of $d_G = 3$ for RNAs with a random nucleotide sequence and the most extreme value $d_G = 5$ for viral RNAs, we obtain $\Delta F(d_G = 5)/\Delta F(d_G = 3) \approx 1.6(\epsilon b/l)^{-1/4}$. If we consider the upper limit of $\epsilon b/l \approx 1$ for (8) to be valid, we get $\Delta F(d_G = 5)/\Delta F(d_G = 3) \approx 1.6$. In other words, the free-energy gain for a chain with $d_G = 5$ is more than 60% larger than that for a chain with $d_G = 3$ (if we ignore the unknown numerical prefactors). Note that because $R_c^2 l/v$ is on the order of 100, the free-energy enhancement can be substantial.

Finally, we consider the encapsulation of randomly branched polymers for both quenched and annealed cases. As we argued in the introduction, the statistical behavior of the annealed ones seems closer to single-stranded RNAs than that of the quenched ones. Even though $d_G = 4$ for both types of branched polymer, their fractal dimensions differ under good and poor solvent conditions [46]. There are a few interesting conclusions we can draw from our calculations. First, we find that the optimal packaging of annealed RNAs are larger than that of quenched ones, $M_*^{\text{ann}}/M_*^{\text{que}} = 35/27 \approx 1.30$, in the regime where the second virial approximation is valid, i.e., $M_* v/R_c^2 D_* \ll 1$. Here, the subscripts *ann* and *que* refer to the annealed and quenched degrees of disorder, respectively. This would imply that 30% more polynucleotide would be absorbed if RNA is allowed to restructure its branching. Annealing would therefore strongly increase the ability of the capsid to package genome.

If we consider the adsorbed layer thickness, however, then we find that $D_*^{\text{ann}}/D_*^{\text{que}} \propto (\epsilon b/l)^{-20/3} \gg 1$ if $\epsilon b/l < 1$. This is not entirely surprising, because (i) a longer annealed polymer is encapsulated than a quenched one for the same interaction strength and (ii) a self-avoiding annealed polymer in free solution is larger than a quenched one if their length is the same (see the introductory section). A more careful analysis shows, in fact, that unless $\epsilon b/l \gtrsim (l/R_c)^{1/7}$ the optimal annealed chain fills the entire cavity of the capsid, i.e., $R_c = D_*$. Interestingly, something similar was found in the more elaborate calculation of Lee and Nguyen [42]. As pointed out in Ref. [42], for some viruses this seems to be the case because the interaction strength between RNA and coat proteins is not strong and as such there is no well-defined adsorption layer.

On the other hand, we do find that the free energy of encapsulation for annealed randomly branched polymers to be less than that of quenched ones, unless $\epsilon b/l > 1$. In that case, the thickness layer $D_* \ll R_c$ and the the RNA sitting on the inner surface of the capsid is so dense that the second-virial approximation that we relied on in our analytical solutions to the governing equations fails. Clearly, a much more detailed study is needed to settle this

issue. We leave for future work a direct comparison of the field-theoretical approaches for quenched and annealed branched polymers in the context of encapsulation.

4 Conclusions

In conclusion, we have shown that if the primary sequence of single-stranded RNA of viruses have indeed evolved to optimize their secondary structure so as to obtain higher levels of branching and hence compactness, as it has recently been suggested (Gopal et al., submitted), this does not necessarily lead to packaging of more RNA in virus capsids. However, according to our calculations, a higher level of branching considerably increases the free-energy gain of packaging of viral RNA rather than other types of RNA. More specifically, the secondary structure of RNA resulting from its branching has a significant impact on the efficiency of genome packaging. In head-to-head competition between viral and other types of RNA in co-assembly with virus coat proteins in infected cells, higher degree of branching could indeed give viral RNAs a competitive edge over other RNAs.

References

1. Chiu, W., Burnett, R.M., Garcea, R.L.: *Structural Biology of Viruses*, eds. Oxford University Press, Oxford (1997)
2. Caspar, D.L.D., Klug, A.: Physical principles in the construction of regular viruses. *Quant. Biol.* **27**, 1–24 (1962)
3. Fraenkel-Conrat, H., Williams, R.C.: Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Nat. Acad. Sci. USA* **41**, 690–698 (1955)
4. McPherson, A.: Micelle formation and crystallization as paradigms for virus assembly. *BioEssays* **27**, 447–458 (2005)
5. Larson, S.B., McPherson, A.: Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.* **11**, 59–65 (2001)
6. Cuillel, M., Berthet-Colominas, C., Timmins, P.A., Zulauf, M.: Reassembly of brome mosaic virus from dissociated virus. *Eur. Biophys. J.* **15**, 169–176 (1987)
7. Bancroft, J.B., Hiebert, E., Rees, M.W., Markham, R.: Properties of cowpea chlorotic mottle virus, its protein and nucleic acid. *Virology* **34**, 224–239 (1968)
8. Bancroft, J.B., Hiebert, E., Bracker, C.E.: The effects of various polyanions on shell formation of some spherical viruses. *Virology* **39**, 924–930 (1969)
9. Hiebert, E., Bancroft, J.B., Bracker, C.E.: The assembly in vitro of some small spherical viruses, hybrid viruses and other nucleoproteins. *Virology* **34**, 492–508 (1968)
10. Zlotnick, A.: To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67 (1994)
11. Zlotnick, A., Aldrich, R., Johnson, J.M., Ceres, P., Young, M.J.: Mechanism of capsid assembly for an icosahedral plant virus. *Virology* **277**, 450–456 (2000)
12. Hu, Y., Zandi, R., Anavitarte, A., Knobler, C.M., Gelbart, W.M.: Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. *Biophys. J.* **94**, 1428–1436 (2008)
13. Ren, Y., Wong, S.-M., Lim, L.-Y.: In vitro reassembled plant virus-like particles for loading of polyacids. *J. Gen. Virol.* **87**, 2749–2754 (2006)
14. Sikkema, F.D., Cornellas-Aragones, M., Fokkink, R.G., Verduin, B.J.M., J.Cornelissen, J.L.M., Nolte, R.J.: Monodisperse polymer-virus hybrid nanoparticles. *Org. Biomol. Chem.* **5**, 54–57 (2007)
15. Tsvetkova, I., Chen, C., Rana, S., Kao, C., Rotello, V., Dragnea, B.: Pathway switching in templated virus-like particle assembly. *Soft Matter* **8**, 4571–4576 (2012)
16. Benjamin, J., Ganser-Pornillos, B.K., Tivol, W.F., Sundquist, W.I., Jensen, G.J.: Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *J. Mol. Biol.* **346**, 577–588 (2005)
17. Briggs, J.A.G., Grunewald, K., Glass, B., Forster, F., Krausslich, H.-G., Fuller, S.D.: The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure (Lond.)* **14**, 15–20 (2006)

18. Ganser, B.K., Li, S., Klishko, V.Y., Finch, J.T., Sundquist, W.I.: Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80–83 (1999)
19. Nguyen, T.T., Bruinsma, R.F., Gelbart, W.M.: Continuum theory of retroviral capsids. *Phys. Rev. Lett.* **96**, 078102 (2006)
20. Nguyen, T.T., Bruinsma, R.F., Gelbart, W.M.: Elasticity theory and shape transitions of viral shells. *Phys. Rev. E* **72**, 051923 (2005)
21. Yul, Z., Dobro, M.J., Woodward, C.L., Levandovsky, A., Danielson, C.M., Sandrin, V., Shi, J., Aiken, C., Zandi, R., Hope, T.J., Jensen, G.J.: Unclosed HIV-1 capsids suggest a curled sheet model of assembly. *J. Mol. Biol.* **425**, 112–123 (2013)
22. Hicks, S.D., Henley, C.L.: Irreversible growth model for virus capsid assembly. *Phys. Rev. E* **74**, 031912 (2006)
23. Levandovsky, A., Zandi, R.: Nonequilibrium assembly, retroviruses and conical shape. *Phys. Rev. Lett.* **102**, 198102 (2009)
24. Grime, J.M.A., Voth, G.A.: Early stages of the HIV-1 capsid protein lattice formation. *Biophys. J.* **103**, 1774–1783 (2012)
25. Bruinsma, R.F., Gelbart, W.M., Reguera, D., Rudnick, J., Zandi, R.: Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* **90**, 248101 (2003)
26. Borodavka, A., Tuma, R., Stockley, P.G.: Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. USA* **109**, 15769–15774 (2012)
27. Ni, P., Wang, Z., Ma, X., Das, N.C., Sokol, P., Chiu, W., Dragnea, B., Hagan, M., Kao, C.C.: An examination of the electrostatic interactions between the N-terminal tail of the brome mosaic virus coat protein and encapsidated RNAs. *J. Mol. Biol.* **419**, 284–300 (2012)
28. Hagan, M.F.: A theory for viral capsid assembly around electrostatic cores. *J. Chem. Phys.* **130**, 114902 (2009)
29. van der Schoot, P., Bruinsma, R.: Electrostatics of an RNA virus. *Phys. Rev. E* **70**, 061928 (2005)
30. Belyi, V.A., Muthukumar, M.: Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. USA* **103**, 17174–17178 (2006)
31. Siber, A., Zandi, R., Podgornik, R.: Thermodynamics of nanospheres encapsulated in virus capsids. *Phys. Rev. E* **81**, 051919 (2010)
32. Prinsen, P., van der Schoot, P., Gelbart, W.M., Knobler, C.M.: Multishell structures of virus coat proteins. *J. Phys. Chem. B* **114**, 5522–5533 (2010)
33. Hu, T., Zhang, R., Shklovskii, B.I.: Electrostatic theory of viral self-assembly. *Physica A* **387**, 3059–3064 (2008)
34. Ting, C.L., Wu, J., Wang, Z.-G.: Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proc. Natl. Acad. Sci. USA* **108**, 16985–16990 (2011)
35. van der Schoot, P., Zandi, R.: Kinetic theory of virus capsid assembly. *Phys. Biol.* **4**, 296–304 (2007)
36. Zandi, R., van der Schoot, P.: Size regulation of ss-RNA viruses. *Biophys. J.* **96**, 9–20 (2009)
37. Šiber, A., Lošdorfer Božić, A., Podgornik, R.: Energies and pressures in viruses: contribution of non-specific electrostatic interactions. *Phys. Chem. Chem. Phys.* **14**, 3746–3765 (2012)
38. Šiber, A., Podgornik, R.: Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Phys. Rev. E* **78**, 051915 (2008)
39. Lin, H., van der Schoot, P., Zandi, R.: Impact of charge variation on the encapsulation of nanoparticles by virus coat proteins. *Phys. Biol.* **9**, 066004 (2012)
40. Dobrynin, A.V., Rubinstein, M.: Theory of polyelectrolytes in solutions and at surfaces. *Prog. Polym. Sci.* **30**, 1049–1118 (2005)
41. Elrad, O.M., Hagan, M.F.: Encapsulation of a polymer by an icosahedral virus. *Phys. Biol.* **7**, 045003 (2010)
42. Lee, S.I., Nguyen, T.T.: Radial distribution of RNA genomes packaged inside spherical viruses. *Phys. Rev. Lett.* **100**, 198102 (2008)
43. De Gennes, P.-G.: Statistics of branching and hairpin helices for the dAT copolymer. *Biopolymers* **6**, 715–729 (1968)
44. Grosberg, A., Gutin, A., Shakhnovich, E.: conformational entropy of a branched polymer. *Macromolecules* **28**, 3718–3727 (1995)
45. Gutin, A.M., Grosberg, A.Y., Shakhnovich, E.I.: Polymers with annealed and quenched branchings belong to different universality classes. *Macromolecules* **26**, 1293–1295 (1993)
46. Grosberg, A.Y.: Disordered polymers. *Phys. Uspekhi* **40**, 125–158 (1997)
47. Bundschuh, R., Hwa, T.: Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E* **65**, 031903, 1–22 (2002)
48. Gopal, A., Zhou, Z.H., Knobler, C.M., Gelbart, W.M.: Visualizing large RNA molecules in solution. *RNA* **18**, 284–299 (2012)

49. Yoffe, A.M., Prinsen, P., Gopal, A., Knobler, C.M., Gelbart, W.M., Ben-Shaul, A.: Predicting the sizes of large RNA molecules. *Biophys. J.* **105**, 16153–16158 (2008)
50. Fang, L.T., Gelbart, W.M., Ben-Shaul, A.: The size of RNA as an ideal branched polymer. *J. Chem. Phys.* **135**, 155105 (2011)
51. Borisov, O.V., Vilgis, T.A.: Polyelectrolyte manifolds. *Europhys. Lett.* **35**, 327–333 (1996)
52. Schwab, D., Bruinsma, R.: Flory theory of the folding of designed RNA molecules. *J. Phys. Chem. B* **113**, 3880–3893 (2009)
53. Yaman, K., Pincus, P., Solis, F., Witten, T.A.: Polymers in curved boxes. *Macromolecules* **30**, 1173–1178 (1997)
54. van der Spoel, D., Feenstra, K.A., Hemminga, M.A., Berendsen, H.J.C.: Molecular modeling of the RNA binding N-terminal part of Cowpea chlorotic mottle virus coat protein in solution with phosphate ions. *Biophys. J.* **71**, 2920–2932 (1996)
55. Moghaddam, S., Caliskan, G., Chauhan, S., Hyeon, C., Briber, R.M., Thirumalai, D., Woodson, S.A.: Metal ion dependence of cooperative collapse transitions in RNA. *J. Mol. Biol.* **393**, 753–764 (2009)