THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Role of RNA Branchedness in the Competition for Viral Capsid Proteins

Surendra W. Singaram,[†,∥] Rees F. Garmann,[†,⊥] Charles M. Knobler,[†] William M. Gelbart,[†,‡,§] and Avinoam Ben-Shaul*[,∥]

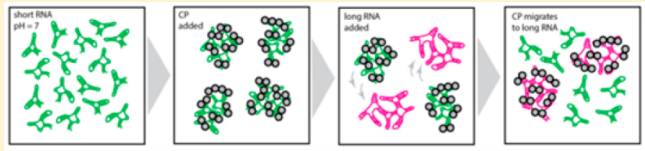
[†]Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90095, United States

[‡]California NanoSystems Institute and [§]Molecular Biology Institute, UCLA, Los Angeles, California 90095, United States

[∥]Institute of Chemistry and the Fritz Haber Research Center, The Hebrew University, Jerusalem, 91904 Israel

Supporting Information

ABSTRACT: To optimize binding—and packaging—by their capsid proteins (CP), single-stranded (ss) RNA viral genomes often have local secondary/tertiary structures with high CP affinity, with these "packaging signals" serving as heterogeneous nucleation sites for the formation of capsids. Under typical *in vitro* self-assembly conditions, however, and in particular for the case of many ssRNA viruses whose CP have cationic N-termini, the adsorption of CP by RNA is nonspecific because the CP concentration exceeds the largest dissociation constant for CP−RNA binding. Consequently, the RNA is saturated by bound protein before lateral interactions between CP drive the homogeneous nucleation of capsids. But, before capsids are formed, the binding of protein remains reversible and introduction of another RNA species—with a different length and/or sequence—is found experimentally to result in significant redistribution of protein. Here we argue that, for a given RNA mass, the sequence with the highest affinity for protein is the one with the most compact secondary structure arising from self-complementarity; similarly, a long RNA steals protein from an equal mass of shorter ones. In both cases, it is the lateral attractions between bound proteins that determines the relative CP affinities of the RNA templates, even though the individual binding sites are identical. We demonstrate this with Monte Carlo simulations, generalizing the Rosenbluth method for excluded-volume polymers to include branching of the polymers and their reversible binding by protein.



1. INTRODUCTION

One of the remarkable characteristics of single-stranded (ss) RNA viruses is that many of them can self-assemble *in vitro* from purified RNA and capsid protein components. This was first demonstrated in 1955 by Fraenkel-Conrat and Williams,[1] who reported the reconstitution of infectious tobacco mosaic virus (TMV) particles—each consisting of a single 6400-nucleotide (nt)-long ssRNA genome protected by a hollow cylinder made up of 2130 copies of its 159-residue coat protein. Concerted studies over the following decades established that the nucleation of the cylindrical capsid is initiated by selective binding of coat proteins to a specific stem-loop in the secondary structure of the viral RNA. Insertion of this nucleotide sequence into an arbitrary RNA molecule results in its efficient encapsidation by TMV coat protein into monodisperse rods whose length is determined by the length of the RNA.

In 1967, a second example of in vitro virus self-assembly from purified components was provided by Bancroft and Hiebert,[2] who showed that a *spherical* virus—cowpea chlorotic mottle virus (CCMV)—could be reconstituted in this way. Subsequent work by Bancroft and co-workers[3] established that the CCMV coat protein (CP) was similarly capable of packaging heterologous ssRNA from other viruses and nonviral ssRNA, as well as flexible anionic synthetic polymers, into capsids identical in size to the wildtype virus, i.e., 28 nm-diameter shells consisting of 180 copies of the CP. Work by Zlotnick et al.[4] has explored substoichiometric CP−RNA intermediates and their role in determining nucleation pathways for formation of complete capsids. More recently, we have shown how the strong preference of CCMV CP for 28 nm-diameter shells leads to the formation of multiplets when CP is added to ssRNA of increasing length: for RNA twice as long as the ≈3000nt-long CCMV RNA, pairs (doublets) of capsids are involved in the packaging of RNA, while for RNAs three and four times longer triplets and quadruplets are formed.[5]

Further, it has been demonstrated for CCMV[5−7] that the strength of the lateral interactions between CP responsible for capsid formation from RNA-bound CP can be controlled by solution pH. Specifically, the strength of CP−CP attraction can increase upon lowering the pH. While RNA binding sites are completely saturated[8] upon mixing RNA and CP at neutral pH and low ionic strength, lowering the pH to 6 or lower is necessary to form 180-CP capsids that are capable of protecting the RNA against nucleases. Indeed, at low pH and *high* ionic

Received: July 5, 2015
Revised: August 21, 2015
Published: October 4, 2015

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

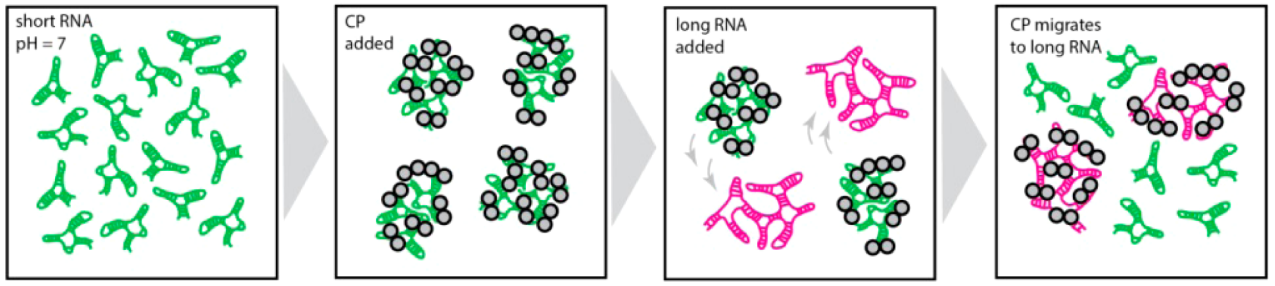


Figure 1. Schematic illustration of competition between short and long RNAs for binding of capsid protein. From left to right: Short RNAs are initially saturated with capsid proteins, but upon the addition of long RNAs all the proteins migrate to the longer RNAs.

strength, capsids form in the *absence* of RNA. In addition, these studies have shown that the binding of CP to RNA is *reversible* at neutral pH, but not at the lower pH where effective CP–RNA binding affinities are strongly enhanced by lateral interactions between bound CP. This effect is seen most dramatically in experiments in which two different RNA molecules are made to compete against one another for an amount of CP insufficient for packaging both.[6]

More explicitly, when an RNA of arbitrary sequence and length (e.g., 3000nt) is incubated at neutral pH with just enough CCMV CP to completely saturate it, all of the CP is found to be bound to the RNA. (Note that, because of the 10 cationic residues per N-terminus, saturation of the RNA implies one CP per 10nt of RNA, corresponding to a CP:RNA mass ratio of 6:1.) Lowering the pH to a value below 6 then results in complete packaging of the RNA into RNase-resistant capsids. Similarly, if a shorter RNA (say, 1000nt) is subjected to the same protocol, it too will bind all the CP at neutral pH and be completely packaged upon lowering of the pH. If, on the other hand, equal masses of the two RNA molecules are incubated together with CP at a CP:total RNA mass ratio of 3, so that there is insufficient CP to package all of the RNA in the mixture, all of the protein will be bound at neutral pH by the longer RNA (and none by the shorter) and only the longer will be packaged into protective capsids upon pH lowering.[6] Still more dramatically, if the shorter RNA is incubated alone with the CP at neutral pH and CP:RNA = 6, *followed by addition of and incubation with an equal mass of the longer RNA*, pH lowering leads to the longer RNA being exclusively packaged and the short RNA "stripped" of its protein, despite the longer RNA having been added later to the solution: see Figure 1. Alternatively, if the longer molecule is incubated first with CP it retains all of the protein after addition of the shorter RNA, and is the only molecule packaged upon pH lowering. From these facts it is clear that the order of incubation at neutral pH, where the CP binding is reversible, is not important.

In this paper, we argue that competition among different RNA molecules for viral capsid protein is determined by the differing extents to which bound proteins are able to interact laterally with one another. In particular, for molecules of the same length (hence, with the same number of nucleotides, and CP binding sites), we show that the best competitor is the RNA that is made most compact by its sequence-dependent secondary structure. For molecules of different length but comparable degrees of effective branching due to secondary structure formation, the longer one wins because it allows protein to "condense"—satisfy its attractive lateral interactions—with a smaller "surface-to-volume" ratio. These phenomena are examples of "specificity" (i.e., the preference of CP for one RNA over another) and are offered as complements to the competitive CP binding effects provided by local "packaging signals".[9−11] By using a common CP affinity (energy lowering) for all the RNA binding sites in all of the molecules (linear, branched, compact, and extended) in our model, we are able to isolate and highlight the effects due to lateral interactions of the bound proteins.

2. THEORY

A simple way to "level the playing field" for competition between two or more molecules for the binding of protein is to have equal masses of each competitor, so that they present equal numbers of binding sites, and a limited amount of protein. As already mentioned in the Introduction, the viral CP–RNA binding experiments that motivate our work typically involve two RNA molecules of identical length (i.e., the same number, N, of nt) but different sequences and hence different secondary structures. Alternatively, they may involve RNAs of length N competing with twice as many RNAs of length $N/2$. The coarse-grained properties of the ensemble of secondary structures with which we will be concerned are the overall size (radius of gyration) and the nature of the branching that result from these structures, in particular the distribution of the orders and the positions of the branch points. The "branch points"—of third- and fourth-order, for example—are associated with single-stranded loops from which three or four duplexes emanate. The ssRNA molecules in these experiments are long (viral length)—comprised of a few thousand nt, and are capable of binding hundreds of capsid proteins. Thus, fluctuations in the distribution of CP between the competing species are quite small, and the experiments can therefore be modeled by focusing on just one pair of different RNAs competing for a given total amount of CP.

2.1. Model. To simulate the competition for binding of CP between long vs short RNAs, or branched vs linear RNAs, or compact vs extended branched RNA molecules, we use the simplest model that captures the essential qualitative aspects of this phenomenon. A basic premise of the model is that CP binding does not affect the secondary structure of the RNA molecule. On the other hand, attractions between proteins bound to nearest-neighbor sites will be a dominant factor in determining the tertiary structure of the RNA, i.e., its configuration in 3D space. As in several previous studies[12−15] the secondary structures of the ssRNA molecules will be mapped onto their tree graph representations, whereby base-pair (bp) duplexes are treated as rigid edges (all of the same length) and the single-stranded loops (the tree vertices) connecting them are regarded as flexible joints. The basic unit in the branched polymer is a duplex-stem (edge) and its

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

attendant ss-loop (vertex). For computational reasons the largest tree graphs considered in this work comprise 50 stem-loop pairs, corresponding to RNA chains of about 1000 nt in which about 60% of the nt are typically paired in duplexes whose average length is about 5bp.

When we compete one RNA molecule against another of a different length we attribute to them the same branchedness, i.e., the same relative numbers of 1-fold vertices (hairpin-loops), of 2-fold vertices (connecting only two duplexes), and of third- and higher- order branch points. In this way we can focus on the effect of different numbers of binding sites on the ability of a molecule to compete for capsid proteins. In the same vein, when we compete *equal*-length RNA molecules, we either attribute different distributions of vertex orders to them (e.g., as in the case of a branched vs linear RNA) or we keep the vertex-order distributions the same but scramble the vertices so that the molecules are more extended or more compact.

Finally, the *tertiary* structures of the tree graphs will be represented by embedding them with different configurations on a two-dimensional (2D) square lattice. While motivated by computational simplicity, this limitation to 2D structures is not unreasonable considering that the RNA backbone of viral capsids serves as the template for the nucleation of a 2D (albeit curved) protein shell protecting the genomic material. The use of a square lattice, where the maximal vertex order is 4, is not a severe restriction, in view of the fact that fifth- or higher-order vertices in RNA secondary structures are very rare.[16,17] Accordingly, in translating the original RNA sequences to tree graphs we have counted all the loops of order five or larger as fourth-order vertices. We note that 4 is also the number of contacts per dimer in the 180-CP capsids of CCMV.

In aqueous solution, over a broad range of pH and ionic strength conditions (including physiological), the CP of CCMV exist as dimers, serving as the fundamental assembly units of the viral protein shell.[18] Each of the CP-dimer building blocks is attracted nonspecifically to the negatively charged RNA genome through the two cationic N-terminal arms of its constituent monomers, totaling 20 positive charges. This number, 20, is also the average number of nt negative charges per stem-loop pair because, on average, the RNA duplexes consist of 5bp and ss-loops typically contain 10nt.[15] Furthermore, the physical size ("footprint") of a stem-loop pair is also comparable to that of a CP-dimer. Thus, at full coverage ("saturation") the number of CP dimers bound to an RNA molecule equals its number of stem-loop pairs: each vertex-edge pair in our tree-graph lattice model is a potential site for the reversible binding of one CP dimer (hereafter simply CP), as illustrated in Figure 2.

Our model assumes attractive CP−CP interactions between CP pairs occupying nearest-neighbor (NN) sites, whether bonded by a stem (see, e.g., vertices 1 and 2 in Figure 2b) or not (e.g., vertices 1 and 4). On energetic grounds these attractive interactions obviously favor compact conformations of the CP-dressed branched polymer, which on the other hand are generally disfavored entropically. In our Monte Carlo (MC) simulations of the CP-dressed polymers, we allow for conformational changes of the branched polymer, as well as for rearrangements of the reversibly bound CP on the branched tree backbone, enabling the structure to reach thermodynamic equilibrium. Only self-avoiding polymer conformations are allowed, thus respecting excluded-volume interaction.

Note that, as discussed above, we explicitly allow for changes in the *tertiary* structure of the tree-graph representations of the RNA, due to lateral interactions between the particles bound to them. But the topology—*connectivity*—of the tree graph does *not* change, even as it is configured differently on the lattice in which it is embedded (see Figure 2). Recalling how the tree-graph connectivity is determined directly from the naked RNA secondary structure, we are effectively neglecting changes in secondary structure due to interaction of the RNA with its capsid protein. A recent study of satellite tobacco mosaic virus,[19] for example, suggests significant differences between the secondary structure of its naked RNA and that of the genome in its mature nucleocapsid. In the present work, on the other hand, where we treat the initial binding of capsid protein by RNA—instead of the formation of a complete viral nucleocapsid—we proceed with the simplifying assumption that the dominant effect of capsid protein is to impose tertiary organization on the RNA secondary structure.

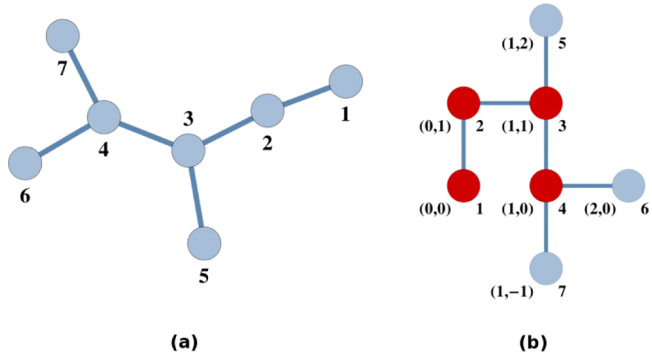


Figure 2. (a) Tree graph corresponding to the secondary structure of a small RNA molecule, with edges and vertices representing base-pair (bp) duplex stems and single-stranded (ss) loops, respectively. (b) Tertiary configuration of this tree graph, now with bound CP (red circles), on a 2D square lattice. The specified (x,y) coordinates define a particular tertiary configuration. With ε (<0), the attraction energy between nearest-neighbor CP pairs, the energy of this particular configuration is 4ε.

Our statistical thermodynamic simulations of CP−RNA binding patterns and competition experiments include the following cases:

(1) One long RNA molecule, represented by a 50-vertex tree graph, competing for 50 CP units against 2 shorter, 25-vertex, molecules. The long and the short tree graphs associated with these molecules each have the same vertex order distribution, corresponding (as explained below in section 2.2) to that of random RNA sequences with uniform nt composition (i.e., equal numbers of A, U, G, and C).

(2) Two 50-vertex molecules sharing 50 CP and—separately—two 25-vertex molecules sharing 25 CP. Here the idea is to investigate the possibility of unequal binding of CP by identical branched polymers. For these studies we use again the branching pattern associated with random sequences and uniform nt composition.

(3) A compact 50-vertex polymer competing with an extended 50-vertex polymer for 50 CP. The vertex-order distributions of the compact and extended trees are identical, but their radii of gyration are markedly different. The procedure for generating these compact and extended trees, involving the repositioning of branch points within a tree graph, is outlined in section 2.2.

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

(4) The same compact, 50-vertex, branched polymer as in point 3, competing for 50 CP against a 50-vertex *linear* polymer.

2.2. Generating Conformations of CP-Dressed Tree Graphs. We have used two complementary procedures to simulate the competition experiments, hereafter referred to as *thermodynamic* and *kinetic* simulations, respectively. In both approaches, in analogy to a previous extension[20,21] of the Rosenbluth Monte Carlo (RMC) algorithm,[22,23] we first generate representative ensembles of *low free energy configurations* of the relevant *CP-dressed polymer* − e.g., 1000 conformations of the 50-vertex tree, dressed with M CP. For each of these ensembles (i.e., for each value of M between 0 and 50) we calculate properties of the dressed polymer, such as its radius of gyration, as well as its partition function and thus any desired thermodynamic function. In the thermodynamic simulations, we use these partition functions to evaluate the probability of any division of the $M = M_1 + M_2$ proteins between the competing polymers (e.g., the linear vs branched polymers), thus obtaining the equilibrium distribution of bound proteins and the winner of the competition, if any. The kinetic simulations employ a variant of the Metropolis algorithm[23,24] whereby CP are exchanged between the competing polymers, allowing for concomitant changes in the spatial conformations of both polymers. The configurations of the competing polymers are sampled from the ensembles of configurations generated by our RMC procedure. To satisfy the principle of detailed balance the exchange probabilities in these simulations are weighted according to the joint partition functions of the competing structures before and after the exchange, rather than simply their Boltzmann weights. Below we outline our RMC procedure for generating CP-dressed polymers. Their use in the thermodynamic and kinetic approaches is described in sections 2.3 and 2.4, respectively.

Our goal is to generate an ensemble of conformations of a polymer with an arbitrary branching pattern comprised of L vertices with M CP bound to its backbone, with $f = M/L$ denoting the overall fraction of occupied vertices. Using the branched polymer in Figure 2 as an illustrative example, we first randomly choose one of its vertices (no matter of what order), e.g., vertex 1, place it at site $x,y = 0,0$ of the 2D square lattice, and with probability $f_1 = f = M/L$ populate this site with one CP. Next, in order to begin generating ("growing") an L-vertex/M-protein chain conformation, we randomly pick one of its bonded vertices, which in this example must be vertex 2, and note that there are now 8 possible states for it, corresponding to placing this vertex in any of the 4 vacant neighboring sites, in each case with or without a bound CP. The choice illustrated in Figure 2, where vertex 2 is at $x,y = 0,1$ with a bound CP, is sampled with probability $\alpha f_2/w_2$ where $f_2 = (M - 1)/(L - 1)$, and $\alpha = \exp(-\varepsilon/kT)$ with T the temperature and k Boltzmann's constant. The sum $w_2 = 4f_2\alpha + 4(1 - f_2)$ is conveniently referred to as the local partition function of vertex 2. We similarly proceed to the 3-fold coordinated vertex 3, which is further connected to two additional bonds leading to vertices 4 and 5. Again we randomly connect one of those to the partially formed tree; note, however, that if vertex 4 is chosen first then both vertices 3 and 4 are still not "saturated" and the next vertex (vertex 5, 6, or 7) may be connected to either of them; the choice is arbitrary. This procedure continues until all bonds are satisfied.

In more general terms, after placing $i - 1$ vertices of an arbitrary tree graph in positions $\mathbf{r}_1,\mathbf{r}_2,...,\mathbf{r}_{i-1}$ there is always at least one unsaturated vertex (and when $i - 1 = L - 1$ there is only one such vertex). We add the next vertex by randomly choosing one of the unsaturated vertices, e.g., vertex k, whose lattice position is $\mathbf{r}_k$, and randomly pick one of its four neighboring sites $\mathbf{r}_k \pm \mathbf{e}_x$ or $\mathbf{r}_k \pm \mathbf{e}_y$, e.g., $\mathbf{r}_k + \mathbf{e}_x$ ($\mathbf{e}_x$ is a unit vector along the x axis, etc.). Let $p_i(\mathbf{r}_k + \mathbf{u})$ denote the joint probability of placing vertex i at $\mathbf{r}_k + \mathbf{u}$ (where $\mathbf{u} = \pm \mathbf{e}_x, \pm \mathbf{e}_y$) and finding it occupied by a CP, and let $q_i(\mathbf{r}_k + \mathbf{u})$ denote the joint probability of choosing the same site but leaving it empty. These probabilities are given by

$$p_i(\mathbf{r}_k + \mathbf{u}) = \frac{\theta(\mathbf{r}_k + \mathbf{u}) f_i \alpha^{n(\mathbf{r}_k+\mathbf{u})}}{w_i} \tag{1}$$

and

$$q_i(\mathbf{r}_k + \mathbf{u}) = \frac{\theta(\mathbf{r}_k + \mathbf{u})(1 - f_i)}{w_i} \tag{2}$$

Here $f_i = (M - M_{i-1})/(L - i + 1)$ with M_{i-1} denoting the number of CP already bound to the partially generated tree of $i - 1$ vertices and $\theta(\mathbf{r}_k + \mathbf{u}) = 1$ or 0 depending on whether site $\mathbf{r}_k + \mathbf{u}$ is available for accommodating vertex i or is already taken by a previous vertex, respectively. Finally, $n(\mathbf{r}_k + \mathbf{u})$ denotes the number of CP occupying nearest neighbor sites around $\mathbf{r}_k + \mathbf{u}$, the prospective site of vertex i. If the chosen site is already occupied we keep searching for a vacant one; once found we set $\mathbf{r}_i = \mathbf{r}_k + \mathbf{u}$ and continue to place vertex $i + 1$. The local partition function associated with vertex i is

$$w_i = \sum_{\mathbf{u}} \theta(\mathbf{r}_k + \mathbf{u})[f_i \alpha^{n(\mathbf{r}_k+\mathbf{u})} + (1 - f_i)] \tag{3}$$

ensuring the normalization $\sum_{\mathbf{u}}[p_i(\mathbf{r}_k + \mathbf{u}) + q_i(\mathbf{r}_k + \mathbf{u})] = 1$.

Continuing to vertices $i + 1$, ..., L, we end up generating a CP-dressed tree graph configuration ψ specified by the spatial coordinates of the L vertices, $\mathbf{r}_1$, ..., $\mathbf{r}_L$, and their respective CP occupancies. The total CP−CP interaction energy in this conformation is $E(\psi) = n(\psi)\varepsilon$, where $n(\psi) = (1/2)\sum_i n_\psi^{(i)}(\mathbf{r}_i)$ is the total number of NN CP pairs. We disregard here the binding energy of the CP to the polymer backbone because in our simulation of the competition experiments the number of bound CP is fixed−only their *distribution* over the competing polymers is changing.

The overall (often-called) Rosenbluth probability of generating an arbitrary conformation is

$$P_R(\psi) = \frac{\alpha^{n(\psi)} \prod_{i=1}^{L} \eta_i}{W(\psi)} = \frac{M!(L - M)!}{L!} \frac{e^{-E(\psi)/kT}}{W(\psi)} \tag{4}$$

with $\eta_i = f_i$ or $1 - f_i$ denoting the probabilities of finding vertex i populated by a CP or remaining vacant, respectively. The product of these functions does not depend on the order of binding the CP to M of the L vertices, and its numerical value is the inverse of the number of such sequences, i.e., $M!(L - M)!/L!$. Note, however, that this (inverse) binomial factor does not imply that the M CP are randomly distributed over the L vertices; their distribution is dictated by eqs 1 and 2. The denominator in eq 4, $W(\psi)$, is the generalized Rosenbluth factor, defined by[22,23]

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

$$W(\psi) = \prod_{i=1}^{L} w_i(\psi) \tag{5}$$

with w_i given by eq 3

Repeated applications of the procedure above yields the ensemble of conformations of the CP-dressed polymer, $\{\psi\}$, which we use to calculate the averages, $\langle\chi\rangle$, of relevant properties of interest, such as the radius of gyration, R_g, or the CP−CP interaction energy, E. Note, however, that the conformations are sampled according to their RMC probabilities in eq 4, rather than in proportion to their Boltzmann weights in the canonical ensemble $\exp[-E(\psi)/kT] = [L!/M!(L - M)!]W(\psi)P_R(\psi)$. The proper thermodynamic average of a property χ is thus given by

$$\langle\chi\rangle = \frac{\sum_\psi \chi(\psi) e^{-[E(\psi)/kT]}}{\sum_\psi e^{-[E(\psi)/kT]}} = \frac{\sum_\psi P_R(\psi)\chi(\psi)W(\psi)}{\sum_\psi P_R(\psi)W(\psi)} = \frac{\sum_\psi^{(R)} \chi(\psi)W(\psi)}{\sum_\psi^{(R)} W(\psi)} \tag{6}$$

where the sums in the first and second quotients run over all conformations of the polymer considered, whereas those in the third quotient—as emphasized by the superscript (R)— run only over the subensemble of conformations sampled by the RMC algorithm.

Finally, we note that the denominators in the last equation are proportional to the partition function of the system, which in the RMC ensemble of conformations is given by

$$Q(L, M, T) = \sum_\psi e^{-[E(\psi)/kT]} = \frac{L!}{M!(L - M)!} \sum_\psi^{(R)} W(\psi; L, M, T) \tag{7}$$

The numerical value of Q is proportional to the number of configurations sampled. However, this number is irrelevant to our simulations of the competition experiments (i.e., it cancels out) because we are only interested in *ratios* of partition functions corresponding to polymers with the same L and M.

As an illustrative special case of eq 7 we note that for the simple case of a linear tree, with no excluded volume between its segments, and in the total absence of CP−CP interaction—i.e., $\varepsilon = 0$ and hence $w_i = 4$, see eq 3—we find $Q = [L!/M!(L - M)!] \times 4^L$. In this simple case, the (logarithms of the) first and second factors of Q account, respectively, for the translational ("mixing") entropy of the CP on the tree backbone and the conformational entropy of the linear polymer. In the cases of interest in this work these two contributions are nonseparable and rather strongly coupled to each other, as we shall see in section 3.

2.3. Thermodynamic Simulations. Consider a system of two polymers P_1 and P_2 of equal length, L, competing for the binding of M capsid proteins, $(M < L)$. At equilibrium, the probability of finding M_1 proteins on P_1 and $M_2 = M - M_1$ on P_2 is

$$P(M_1; M) = \frac{Q_1(M_1)Q_2(M - M_1)}{Q_{tot}(M)} \tag{8}$$

with $Q_{tot}(M) = \sum_{M_1=0}^{M} Q_1(M_1)Q_2(M - M_1)$ denoting the total partition function of the system in equilibrium. The Helmholtz free energy of the system is thus $A_{eq}(M.L) = -kT \ln Q_{tot}(M, L) \approx -kT \ln[Q_1(M_1^*)Q_2(M - M_1^*)]$, where the second near-equality is based on the maximum term approximation, with M_1^*, $M_2^* = M - M_1^*$ denoting the most probable distribution, namely, the maximal $P(M_1;M)$, reflecting the most probable division of the CP among the two proteins. For very large values of M, the most probable populations, M_1^* and M_2^*, are practically identical to the average equilibrium values $\langle M_1\rangle = M_1^{eq} = \sum_{M_1=0}^{M} M_1 P(M_1;M)$ and $\langle M_2\rangle = M_2^{eq} = \sum_{M_1=0}^{M}(M - M_1) P(M_1; M)$, respectively. For the finite, yet large, values of M in our simulations this is a very good approximation.

We start the competition experiments with an arbitrary initial state where the M proteins are divided between the two species—M_1^i on P_1 and $M_2^i = M - M_1^i$ on P_2—and then let the system relax to the equilibrium value M_1^*, M_2^*. In our thermodynamic simulations we first generate RMC ensembles of (generally about 1000) polymer−CP conformations, for all $L \geq M_1 \geq 0$ and $L \geq M_2 \geq 0$. Using eq 7, we calculate the partition functions $Q_1(M_1)$and $Q_2(M_2)$, and hence the free energy difference between any two states, i and f, corresponding to different divisions of the M proteins between P_1 and P_2:

$$\Delta A = A_f - A_i = -kT \ln \frac{P(M_1^f; M)}{P(M_1^i; M)} \tag{9}$$

In particular, the free energy change from an arbitrary initial state to the equilibrium one is

$$\Delta A = A_{eq} - A_i = kT \ln P(M_1^i; M) \approx -kT \ln \frac{P(M_1^*; M)}{P(M_1^i; M)} \tag{10}$$

Using eq 7, we also calculate the average energies of the two competing species, $\langle E_1(M_1)\rangle$, $\langle E_2(M_2)\rangle$ and hence

$$\Delta E = [\langle E_1(M_1^f)\rangle + \langle E_2(M_2^f)\rangle] - [\langle E_1(M_1^i)\rangle + \langle E_2(M_2^i)\rangle] \tag{11}$$

ΔS can now be derived from $T\Delta S = \Delta E - \Delta A$. Finally, as a measure of the compactness of any tree of interest, naked or CP−dressed, we calculate its average radius of gyration using eq 6 with $\chi(\psi) = R_g(\psi)$ and

$$R_g(\psi) = \sqrt{(1/2L^2)\sum_{i,j}[r_i(\psi) - r_j(\psi)]^2} \tag{12}$$

with $r_i(\psi)$ denoting the position of vertex i in configuration ψ.

The same simulation procedure can be applied to model the exchange of CP between polymers of different length. In our numerical calculations we have only considered the case where a polymer of length L competes for M CP $(0 \leq M \leq L)$ with two polymers of length $L/2$, again generating large ensembles of CP-dressed configurations of both species. Using M_1 to denote the number of proteins on the large polymer, and M_2 and M_3 on the two short polymers, the generalization of eq 8 to this particular case is

$$P(M_1; M) = \frac{Q_1(M_1)\sum_{M_2=0}^{M-M_1} Q_2(M_2)Q_3(M - M_1 - M_2)}{Q_{tot}(M)} \tag{13}$$

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

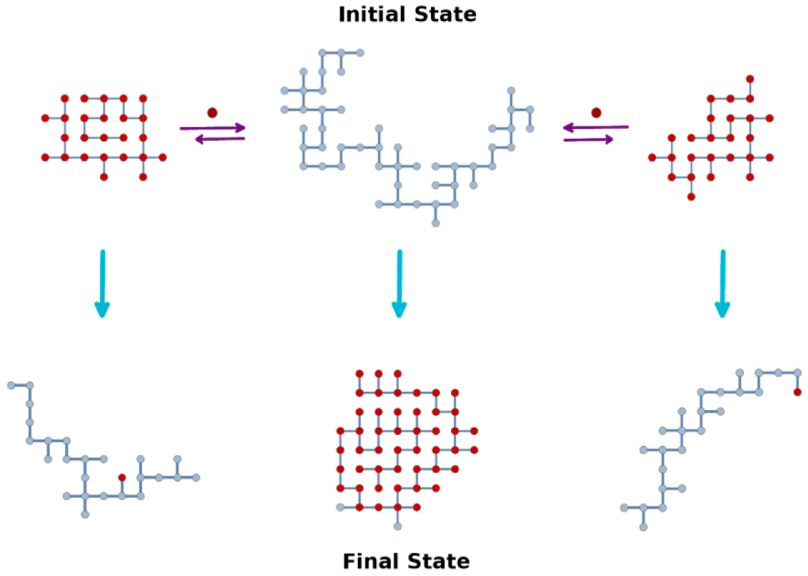


Figure 3. Snapshots from simulation of an experiment where initially all CP are bound to two small trees. Once equilibrium is reached nearly all CP are bound to the large tree.

with $Q_{tot}(M)$ the normalizing factor. All relevant thermodynamic and structural properties can be derived in analogy to the previous case.

2.4. Kinetic Simulations. In this mode of simulation we again begin with an arbitrary initial state with M_1 CP bound to P_1 and M_2 to P_2, both of length L, but now let the system reach equilibrium through CP exchange between the two polymers, coupled to simultaneous conformational changes of the polymers. This procedure resembles the familiar MC simulations involving particle (i.e., CP) exchange except that move probabilities from one state to another are not governed by their relative Boltzmann weights according to the Metropolis criterion. Instead, detailed balance is ensured and equilibrium is reached through attempted moves of CP from one polymer to another, with both the initial and final configurations randomly chosen from the ensembles that we have already generated using the Rosenbluth MC algorithm described in section 2.2. Explicitly, each step begins with a random choice of one of the M CP particles, trying to move it from one tree to the other. If this particle happens to be on P_1, we randomly select one conformation from the previously generated ensemble of conformations of P_1 with M_1-bound CP, and one conformation of P_2 with M_2-bound CP, and attempt a move ending up with another (randomly sampled) conformation of P_1 from the ensemble of conformations of P_1 with $M_1 - 1$ CP and a (randomly sampled) conformation of P_2 from its ensemble of conformations with $M_2 + 1$ CP. Recalling that CP−polymer configurations in the RMC ensembles were not randomly sampled, but rather according to their partition functions, the move is accepted or rejected according to a Metropolis criterion using the partition functions rather than the Boltzmann weight of the configurations involved. In other words, since the probability of any state of the two-polymer system with a given distribution of the CP among them is given by eq 8, the forward and backward rates of transferring one CP from P_1 to P_2 are related by the modified detailed balance ratio

$$\frac{\vec{k}(M_1 \to M_1 - 1)}{\overleftarrow{k}(M_1 - 1 \to M_1)} = \frac{Q_1(M_1 - 1)Q_2(M - M_1 + 1)}{Q_1(M_1)Q_2(M - M_1)} \tag{14}$$

A trial move for transferring a particle from P_1 to P_2 in our kinetic simulations is thus accepted with probability

$$\mathrm{acc}(M_1 \to M_1 - 1) = \min\left[1, \frac{Q_1(M_1 - 1)Q_2(M - M_1 + 1)}{Q_1(M_1)Q_2(M - M_1)}\right] \tag{15}$$

3. RESULTS

To explain the viral assembly competition experiments, the first set of results presented in this section is aimed at understanding why long RNAs steal capsid proteins from shorter RNAs. To this end, in section 3.1, we consider the competition between a 50-vertex tree and two 25-vertex trees. In section 3.2, this computer experiment is contrasted with one where two identical branched RNAs compete with each other. In section 3.3, we consider the competition between two trees with the same vertex-order distributions but with markedly different radii of gyration, i.e., one is significantly more compact than the other. In section 3.4, to further accentuate the role of branchedness in the competition experiments, we treat the competition between a branched tree and a linear tree of the same length. Finally, in section 3.5, a few comments will be made regarding the kinetics of CP redistribution.

Unless specifically stated otherwise, in all simulations we have a CP−CP interaction energy of $\varepsilon = -3kT$, consistent with estimates of the attractive energy between CCMV dimers.[25] For all the competition experiments analyzed below we have carried out both thermodynamic simulations, as well as kinetic—particle-exchange—simulations. The kinetic simulations were sampled every 1000 MC particle exchange steps and generally involved at least 600 000 steps in total, well beyond the time scale needed for the system to reach equilibrium. Temporal evolution profiles of CP populations will thus be shown for shorter time scales. The thermodynamic simulations

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

Table 1. Structure, Energetics, and CP Populations in Competition Experiments[a]

competition P_1 vs P_2	initial → equilibrium CP on P_1 and P_2	$\Delta A/kT$	$\Delta E/kT$	$\Delta S/k$	initial → equilibrium R_g P_1 And P_2
50 vs 2 × 25	0 and 2 × 25 → 48 and 1 × 1 (50)	−29	−27	2	5.0 and 2.1 → 3.3 and 3.0
compact vs extended	0 and 50 → 49 and 1 (50)	−11	2	13	3.2 and 3.2 → 3.0 and 5.2
compact vs linear	0 and 50 → 48 and 2 (50)	−6	11	17	3.2 and 3.0 → 3.0 and 6.2

[a]P_1 and P_2 denote the trees competing for CP. The second column reports the number of CP on the competing trees in the initial (M_1^i, M_2^i, in the first row) and the final equilibrium states ($M_1^{eq} = \langle M_1 \rangle$, $M_2^{eq} = \langle M_2 \rangle$, in the second row). Indicated in parentheses are the most probable CP populations on P_1 at equilibrium, M_1^* (50 in all cases). The third through fifth columns report the changes in free energy, energy, and entropy between the initial and the final equilibrium states. The sixth column tabulates the radius of gyration for each polymer initially and at equilibrium.

are more time-consuming, and are based on ensembles of ∼1000 RMC configurations for each of the tree graphs involved in the competition simulations. From the thermodynamic simulations we present the probability distributions of CP among the competing trees, all showing excellent agreement with the kinetic population profiles. Detailed thermodynamic analyses of energies, free energies and entropies of the competing trees are included in the Supporting Information.

3.1. A Long Tree vs Two Short Trees. In this simulation an $L = 50$ tree (see center of Figure 3) competes for 50 CP with two half-size $L = 25$ trees (depicted on the left and right). The large and small trees are randomly branched polymers, both derived using the same vertex order distribution, as previously determined from analyses of many secondary structures of long random RNA sequences.[17] A pictorial summary of this simulation is presented in Figure 3, showing snapshots from the initial state of the system, where all CP are bound to the two small trees while the large tree is devoid of CP; and the final state where nearly all CP are on P_1 while the two P_2 trees are essentially stripped of all their CP. More precisely, as shown in Table 1 and illustrated in Figure 3, when equilibrium is reached nearly all CP, $M_1 = \langle M_1 \rangle \pm \delta M_1 = 48 \pm 2$ out of 50, end up on the large tree. The temporal changes in the distribution of the CP between the long and short trees is shown in Figure 4, revealing the rapid establishment of the equilibrium distribution and the range of fluctuations around the average CP populations.

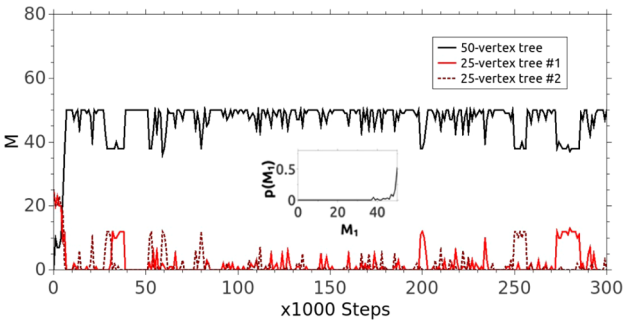


Figure 4. Temporal evolution of the CP populations on the large (black trace) and the two small (red and brown) trees, as obtained by the kinetic (particle exchange) simulations. The CP populations were sampled every 1000 steps. The inset shows the probability distribution of CP on the large tree, as derived from the thermodynamic simulations.

The insert in Figure 4, plotting $P(M_1;M)$ (see eq 13 with $M = 50$), shows the distribution of the number of particles (M_1) bound on the large tree, after the competition between the one 50-vertex and two 25-vertex trees has reached equilibrium. Note that virtually all of the 50 particles are bound to the large tree, even though they *all* started on the two small trees. This effect is also seen, although somewhat less strongly, when the lateral interaction energy between bound particles is weaker than the $(-3kT)$ value used. More explicitly, values of this energy have been estimated[26] for CCMV CP at each of three pHs (4.75, 5.0, and 5.25), and a linear extrapolation of them to neutral pH gives a value of $-1.67kT$. Using this for ε in our simulations, we find a $P(M_1;M)$ distribution that is qualitatively the same as that shown in the Figure 4 insert for $\varepsilon = -3kT$—see Figure S8 in Supporting Information, i.e., the long molecule takes significantly more than its share of the binding particles. We have used the larger value of ε to emphasize more clearly the effect of polymer length on the competition for binding particles. Similarly, we use this value for reporting below the results of competitions between compact and extended branched trees, and between compact/branched and linear trees, where (see Figures S9 and S10) the smaller value of ε again gives qualitatively the same results for the $P(M_1;M)$ particle distributions, i.e., the compact/branched tree binds significantly more than its share of the particles.

The "asymmetry" of the sharing of particles by large and small trees, presenting equal numbers of equal-affinity binding sites, also depends on the ratio of the total number of particles to the total number of binding sites, i.e., on the overall "coverage". For example, the simulations described above (and presented throughout the rest of this work for competition between different kinds of trees) refer to experiments for which the number of capsid proteins is sufficient to saturate each one *or* the other of two competitor RNA species, thus corresponding to an overall coverage of 1/2. Accordingly, the $P(M_1;M)$ distribution in the inset of Figure 4 shows the results for the number of particles on the 50-vertex tree when a total of 50 particles is available to the large and small trees. This is the coverage that gives *maximum* asymmetry of particle sharing: clearly, as the coverage approaches 0 or 1 the asymmetry disappears. But at intermediate values of $^1/_4$ and $^3/_4$, for example, the asymmetry is still quite strong, as shown in Figure S11, where their $P(M_1;M)$ distributions are compared with that for $^1/_2$. Zandi and van der Schoot[27] have emphasized the role of CP:RNA stoichiometry in a related but different context—namely, the crossover from smaller to larger capsid *size* as the overall CP:RNA molar ratio increases (see their Figure 3). In their situation it is the preferred capsid *curvature* (size) that determines the scenarios for stoichiometry dependence, whereas in our case it is the length and branching of the RNA template.

Much as in an Ostwald ripening process, the primary driving force for the transition of nearly all CP from the two small polymers to a single large polymer is the energetic preference of forming one large nearly compact 2D island of CP particles rather than two smaller ones. Indeed, as reported in Table 1 for this particular competition "experiment", the "stealing" of proteins by the large polymer is driven predominantly by

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

energy, with only minor entropy changes. More explicitly, while the relatively high-energy perimeters of both the large and small CP islands are quite ramified, the overall "coastline" of the single large patch is smaller than the combined coastlines of the two smaller CP islands on the small trees. There is, however, one important difference between the present case and familiar ripening processes, such as the coagulation of liquid droplets in 3D[28] or of adsorbate islands on 2D surfaces.[29] Namely, the preferred aggregation of the CP on the larger tree is coupled to a simultaneous change in structure of the embedding substrates, i.e., their host RNA molecules.

In Table 1 we also note a small entropic contribution to the process illustrated in Figure 3. One (minor) contribution to this entropy change is the balance of the conformational entropy loss experienced by the large tree upon its collapse following CP binding, and the concomitant entropy gain by the two small trees. (Note that conformational entropy changes can be significant when polymers of very different branching patterns compete with each other, as will shall see in Secs. 3.3 and 3.4.) Another small entropic contribution is due to the translational ("mixing") entropy of the CP remaining on the small trees and of the vacancies in the large tree.

The establishment of binding equilibria is the result of multiple CP binding-unbinding processes. We note also that the migration of the CP from the two small trees to the large one is reminiscent of a 2D condensation transition. Indeed, were the CP adsorbed on an infinite 2D square lattice, the lateral attraction between neighbors, $\varepsilon = -3kT$, would be strong enough to drive a first-order 2D condensation transition, leading to the coexistence of a dense and a dilute gas phase of CP. In our case the condensed phase resides on the large tree where the CP can form a large island with minimal edge energy, with the few CP on the small trees constituting the dilute phase. We should nevertheless remember that unlike a simple 2D lattice, the branched polymers serving as the substrates of the two phases are finite, highly ramified, and conformationally flexible objects.

3.2. Two Identical Trees. We have also studied the sharing of CP between a pair of identical small trees and between a pair of identical large trees. To this end we have carried out competition simulations involving two 25-vertex trees sharing 25 CP, and—separately—two 50-vertex trees sharing 50 CP. In the latter case, the analogy to a 2D condensation transition is even more compelling than in the previous simulations. The two 50-vertex tree system undergoes rapid phase separation with the majority of CP residing on one tree, enjoying low energy, while the remaining CP form an entropy-rich dilute gas phase on the other tree. As shown in Figure 5, starting with half of the CP population in each of the two 50-vertex trees, symmetry breaking soon takes place with most CP (about 90%) settling on one tree, coexisting with a dilute CP population on the other tree. Although the 50-vertex tree is of finite size and (even in its most compact form) does not provide a perfect square lattice, the CP−CP interaction energy of $\varepsilon = -3kT$—which is substantially stronger than the critical interaction energy for the condensation transition of a 2D lattice gas on a square lattice ($\varepsilon = -1.76kT$)—is strong enough to drive the phase separation of the bound CP, despite the imperfect lattice provided by the underlying tree. We also note that once symmetry is broken, it stays that way on the time scale of our simulations.

The competition between the two shorter, 25-vertex, trees reveals a qualitatively different behavior. Here the imperfection

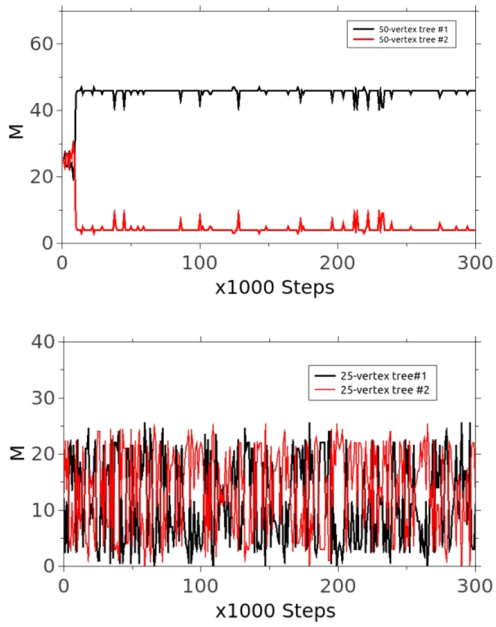


Figure 5. Time dependence of the CP population exchange between two identical trees. Again, populations were sampled every 1000 steps. Top: Starting with the equal sharing of 50 CP among two 50-vertex trees, phase separation occurs rapidly, with the majority of CP condensing on one tree coexisting with a dilute CP population on the other. Bottom: Phase separation also takes place in the two 25-vertex-tree system, but owing to the small system size the CP-population swings often between the two trees.

of the underlying lattices spanned by these short trees, and hence the smaller (average) number of nearest neighbor CP−CP contacts, does not suffice to induce a kinetically stable phase separation. What we see instead is a frequent swinging of the majority of CP between the two trees, with roughly 70% of the CP on one tree and the rest on the other. This behavior is clearly consistent with the rapid migration of CP from the two short trees to the long one in the competition between the 50-vertex tree and the two 25-vertex trees, and will be tested experimentally. The equilibrium distributions for the competition experiments involving two identical trees, as well as simulation snapshots of initial and final CP and tree conformations are shown in the Supporting Information.

3.3. Compact Tree vs Extended Tree. In this section we consider the competition for CP between two branched polymers having the same number of vertices and the same distribution of vertex orders, yet markedly different radii of gyration. Qualitatively, as can be seen in Figure 6, the branch points of the extended polymer reside near its ends, whereas in the compact polymer they are located centrally. More specifically, as in the previous subsections, the vertex-order distributions of both the extended and the compact trees considered are those of tree graphs corresponding to random RNA sequences. However, the compact and extended trees are those with smallest and largest possible R_g, respectively, among the trees with this vertex order distribution. To derive their structure we have first numerically labeled all vertices of an arbitrarily chosen 50-vertex tree having the given vertex-order distribution, and have assigned to it its unique Prüfer sequence.[30] A key property of Prüfer sequences and their one-to-one unique mappings onto branched tree graphs is that *any permutation* ("Prüfer shuffling") *of the sequence*—while generating a new tree graph topology—*leaves the vertex-order*

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

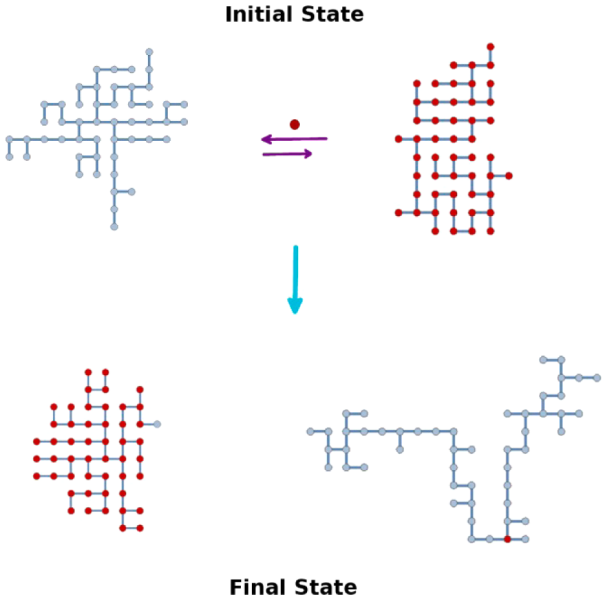


Figure 6. Snapshots from the initial and final equilibrium states obtained in the simulation of the competition between the compact (left) and extended (right) trees. Practically all the CP population is eventually found on the collapsed compact tree.

distribution invariant. Accordingly, we have repeatedly shuffled the starting sequence and for each outcome regenerated the corresponding tree graph and calculated its R_g using eqs 6 and 12. The compact and extended trees used in our simulation represent the trees with smallest and largest radius of gyration obtained in this way.

As indicated in Table 1, the transfer of CP from the extended to the compact tree is driven by the conformational entropy gain of the extended tree, as could be expected in view of its similarity to a linear polymer. From Figure 7 we note a short-lived metastable state (in the time range of ∼10 000−30 000 MCS) on the way to a stable equilibrium (at ∼35 000 MCS) where nearly all CP remain bound to the compact tree, except for small occasional *fluctuations*.

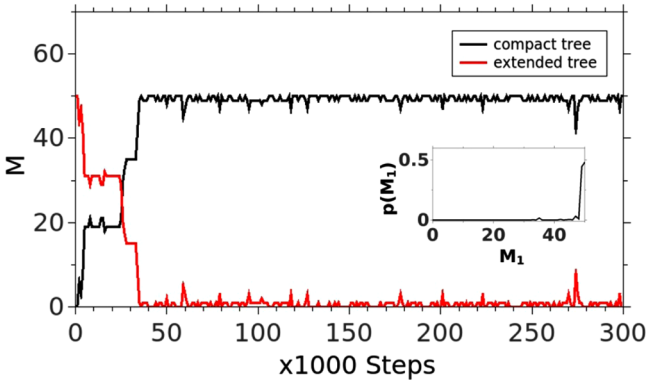


Figure 7. Temporal evolution of the CP sharing between the compact (black trace) and the extended (red) tree. The inset shows the probability distribution of CP on the compact tree, as derived from the thermodynamic simulations.

3.4. Branched Tree vs Linear Tree. In this section we simulate the competition for CP between a linear polymer and a branched polymer of the same length. The main purpose of these simulations is to accentuate the thermodynamic consequences of RNA branchedness with regard to protein binding. The linear polymer may be regarded as representing a nonfolding RNA, e.g., poly-U, for which no secondary structure arises. The branched tree used in the present simulations is identical to the compact tree of the previous section.

The simulations mimic an experiment where L-length branched polymers are added to a solution containing an equal mass of linear polymers that are already saturated with strongly bound CP; no free CP are present in solution. Snapshots from the initial and final states of this simulation, illustrating the outcome of the competition between the two 50-vertex polymers, are shown in Figure 8. The $-3kT$ lateral attractions among the bound CP are strong enough to prompt perfect compaction of the linear polymer in the initial state. However, once the branched polymer is added, the linear polymer gives up essentially all of its CP, despite the energy penalty associated with this move due to the ramified edges of the collapsed branched polymer in comparison to those of the linear polymer. The compensation of this unfavorable energetic contribution is the substantial gain in conformational entropy of the linear polymer, as noted in Table 1. Stated differently, the loss of conformational entropy of the branched polymer upon collapsing to its most compact form is smaller than the corresponding entropy gain by the linear polymer upon giving up its bound proteins.

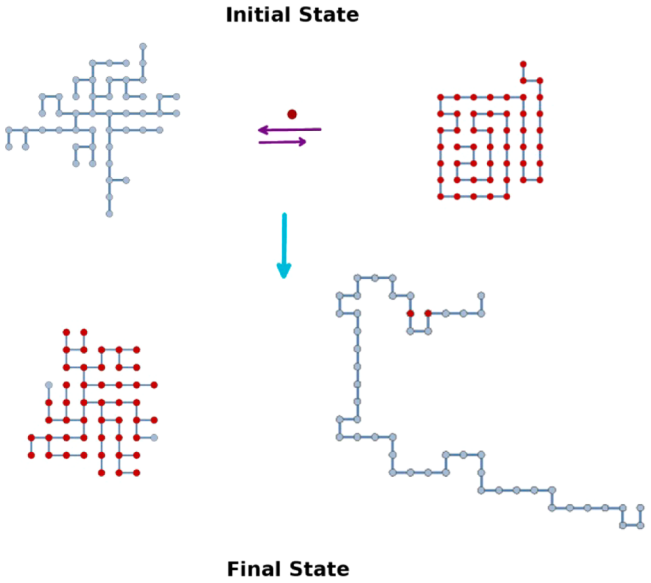


Figure 8. Snapshots from the initial and final equilibrium states obtained in the simulation of the competition between the branched (left) and linear (right) trees. Practically all the CP population is eventually found on the collapsed compact tree. The linear tree enjoys a substantial gain of conformational entropy, overcoming the unfavorable loss of CP−CP energy upon the migration of CP to the branched tree.

The time evolution of the CP populations on the linear vs branched trees is shown in Figure 9, revealing that except for occasional sizable fluctuations, once equilibrium is reached nearly all CP are bound to the branched tree (see inset).

3.5. Kinetics of Protein Exchange between Polymers. There are two possible routes in aqueous solution for CP transfer from one RNA molecule to another. The first is through evaporation−condensation events, whereby CPs desorb from one RNA, diffuse in solution, and eventually

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

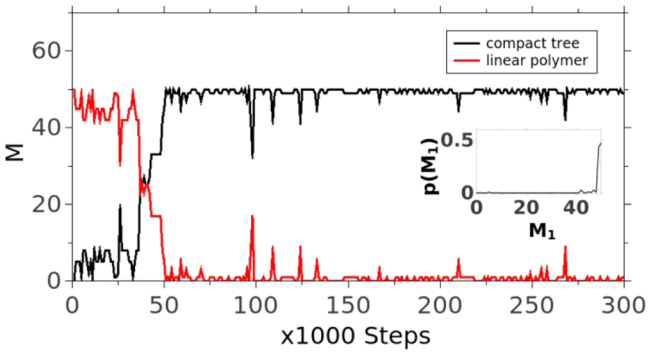


Figure 9. Temporal evolution of the CP populations on the branched (black trace) and the linear (red) trees, as obtained by the kinetic (particle exchange) simulations. The inset shows the probability distribution of CP on the compact tree, as derived from the thermodynamic simulations.

adsorb on another. The second route for CP exchange involves the formation of transient collision complexes through binary encounters of CP-dressed RNA molecules, enabling CP hopping from a binding site on one RNA to a neighboring binding site on another. As in many surface diffusion and chemical reaction events, the unbinding and binding processes take place concertedly, resulting in lower activation energies than those implied by complete unbinding. Recent experimental results[31] suggest that CP exchange through RNA encounters is indeed the more likely mechanism, yet the evaporation−condensation mechanism cannot be ruled out. Notwithstanding this uncertainty, it is very reasonable to assume that the same mechanism is responsible for CP exchange in all the competition experiments modeled in our work. Thus, while we cannot relate the MC step times to real-time events, our kinetic simulations nevertheless shed light on the relative time scales of the various processes involved.

Starting with the competition between the initially naked 50-vertex tree and the two CP-saturated 25-vertex trees, for example, we see from Figure 4 that equilibration is achieved in fewer than 10 000 MC exchange steps. Turning to Figure 5 we are not surprised to observe that the equilibration time for CP sharing among the two 50-vertex trees is significantly slower than that between the two 25-vertex trees. As noted in section 3.2, the stability difference between the small and large trees is even more clearly exhibited by the rapid swinging of the CP population between the two small trees, as compared to the barely fluctuating populations following the symmetry breaking in the competition between the two large trees.

The time to equilibrium in the competition between the compact and extended trees, ∼ 35 000 MC steps (Figure 7), is longer than the time (∼10 000 steps) to equilibrium between two identical branched trees (Figure 5). The origin of this difference is 2-fold. First, the competition between the two identical 50-vertex trees starts with half of the population on each of the two trees whereas in the compact vs extended tree competition all CP are initially adsorbed on the extended (eventually the "loser") tree. Second, when fully saturated with CP, the extended tree collapses to a more compact, and hence more stable, globule than the 50-vertex tree of sections 3.1 and 3.2. The compactness and stability of the initial structure is even more pronounced in the competition between the linear and branched trees, resulting in an even longer time to equilibrium, ≈50 000 MC steps, cf. Figure 9.

4. DISCUSSION

In this work, we have developed a Monte Carlo simulation approach for treating the effects of lateral interactions between proteins on their binding to branched polymers. This problem is motivated by the biologically relevant example of the binding of viral capsid protein to single-stranded RNA molecules, e.g., viral RNA genomes, which behave as effectively branched polymers because of their large extent of secondary structure formation. We have focused here on the competition for binding protein by two or more polymers in order to explain the protein exchange equilibria observed in recent *in vitro* studies of competition between different RNA molecules to be packaged by capsid protein. Our model and simulations are motivated by the disordered complexes of capsid protein bound to RNA in the first, reversible, step of a two-step self-assembly of virus-like particles from RNA and CCMV CP. In particular, in this neutral-pH step the CP−CP interactions are weak enough so that the complexes of bound protein are largely disordered, including only partial portions of spherical nucleocapsid (see, for example, Figure 3b of Garmann et al.'s work[7]). But the underlying physical situation is more general, because it provides basic insights into the ways in which lateral interactions between binding proteins determine the compaction of their polymer substrate.

To perform the simulations reported here we have generalized the Rosenbluth Monte Carlo method of "growing" excluded-volume polymers to include arbitrary branching of the polymer (i.e., arbitrary distributions of vertex orders) as well as reversibly bound proteins that attract each other at short distance and that thereby−through conformational changes of the polymer−are effective at gathering in distant binding sites of the substrate. To elucidate the associated changes in energy (of the protein−protein interactions) and entropy (of the bound proteins and of the polymer conformations), we have treated protein exchange between polymers in the cases of:

(i) equal masses of long and short polymers with the same branching patterns (vertex-order distributions);
(ii) equal masses of equal-length and identical-vertex-order-distribution polymers having very different radii of gyration (because of the different placements of their branch points); and
(iii) equal masses of linear and branched polymers.

In all cases we have suppressed the contribution of "packaging signals" by insisting explicitly that all of the polymers involved are composed of identical binding sites, i.e., proteins bind to them with the same adsorption energy (affinity). As such, our work is complementary to recent theoretical studies in which specific distributions of high-affinity binding sites are shown to facilitate capsid nucleation. In this latter approach,[10,11] polymers with neighboring high-affinity sites−in conjunction with a progressively increasing concentration of binding particles−are shown to be preferentially encapsidated over ones without uniform distributions of binding energies. This role of binding heterogeneity has also been elucidated by molecular dynamics simulations;[32] again the starting point is an explicit assignment of very different protein affinities to the different sites of the polymer.

In our work, on the other hand, all sites are associated with the same protein binding energy. Nevertheless, a distribution of *effective* binding energies arises from the different possibilities for the bound proteins to interact laterally with nearest-neighbor bound proteins−not only ones that are neighbors on

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

the polymer backbone but, more importantly, also ones that occupy distant sites on the polymer. Figure 10, for example,

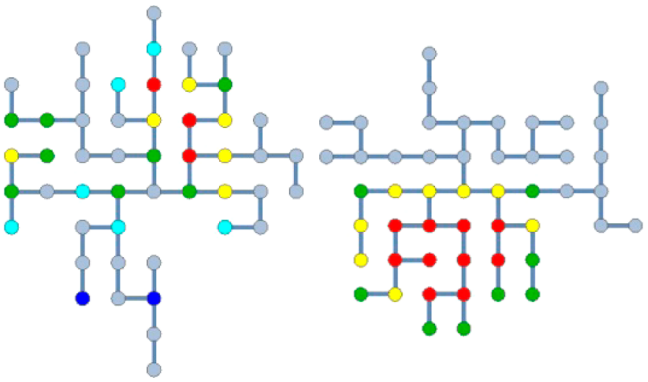


Figure 10. Tertiary configurations of a 50-vertex branched tree at half-coverage, with a typical distribution of bound proteins. The number of occupied nearest neighbors associated with each bound particle are color coded (red corresponding to 4, yellow to 3, and green to 2), reflecting its effective binding energy. Left: the interaction energy between CP on nearest neighbor lattice sites is relatively weak, $\varepsilon = -1kT$. Right: stronger attractive energy, $\varepsilon = -3kT$, results in aggregation of the CPs into one patch.

shows typical configurations of bound particles on an equilibrated 50-vertex branched polymer at half coverage. In the equilibrium ensemble for this coverage every site has a nonzero probability of being occupied. By calculating the average number of occupied nearest-neighbors of each site we can determine the effective energy of each site. Stronger attractive energy results in CP aggregation mediating compaction of the tree they are bound to. Note in particular the localized aggregation of red and yellow particles even in the case of much weaker attractive energy, as in Figure 10, left. In general, it will depend on the nature of the branching pattern of the polymer substrate, e.g., on the vertex-order distribution and the placement of the third and higher-order branch points, and on the strength of CP−CP attraction.

Further work, both experimental and theoretical, will be important for clarifying the relative roles and importance of "packaging signals" and of protein interactions in determining the selective packaging of viral RNA genomes by their capsid protein. In the former case, *local* secondary/tertiary structure of the RNA enhances its affinity for protein, whereas in the latter case it is the extent and nature of branching of the *larger-scale* secondary/tertiary structure that determines the effective binding energies of proteins. These large-scale structures are also important in determining RNA properties—other than binding of protein—relevant to its packaging in viral capsids and virus-like particles. In particular, simulation[33] and theory[34] studies have shown how polymer branching enhances packagability—confinement in small volumes—by reducing the conformational entropy loss and enhancing the interaction of polymer with the capsid interior.

Note that any branched structure, such as the trees representing RNAs, can be made more compact or extended by moving its high-fold junctions closer or father away from its center. For example, Tomato bushy stunt virus (TBSV)[35] and satellite tobacco mosaic virus (STMV)[36−38] were found by SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) studies to have compact and extended secondary structures, respectively, for precisely this reason. More explicitly, SHAPE analyses showed that TBSV has high-fold junctions concentrated near the center of its secondary structure, while STMV has high-fold junctions near its ends. The work by Wu et al.[35] suggests further that the branchedness of the secondary structure correlates with the long-range intragenomic base-pairing interactions that are known to be important in ensuring many genome functions.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b06445.

Additional details concerning the structural and thermodynamic calculations leading to Table 1. More explicitly, free energy, energy, and entropy differences were calculated for every simulated competition experiment as a function of the distribution of the bound CP between the competing trees, as well as the free energy, energy and entropy contributions of each of the trees involved. We also show there how the radius of gyration of each of the trees in question varies with the number of its bound CP. Also shown are snapshots from the simulations of the competition between two identical trees (i.e., two 50-vertex and two-25-vertex trees) along with their equilibrium probability distributions. Finally, we present the results of simulations for which a smaller value ($\varepsilon = -1.67kT$) of the CP−CP interaction energy is used, as well as competition results for smaller (0.25) and larger (0.75) values of the "coverage", i.e., the ratio of (total number of binding particles) to (total number of binding sites) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*(A.B.-S.) E-mail: abs@fh.huji.ac.il.

Present Address

⊥(R.F.G.) Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Mauricio Comas-Garcia for many helpful discussions on the competition of different RNA molecules for capsid protein and Dr. Ajay Gopal for introducing us to, and for teaching us the use and elegance of, the Prüfer sequence and shuffling method. S.W.S. would like to thank Dr. Max Kopelvich and Dr. Esti Livshits for their technical support at UCLA and the Hebrew University, respectively. We would also like to acknowledge the reviewers of our manuscript for several helpful comments and suggestions. A.B.-S. is a member of the Fritz-Haber Research Center and would like to thank the Israel Science Foundation (Grant Number 1448/10) for financial support. C.M.K. and W.M.G. acknowledge support from NSF Grant No. CHE 1051507.

■ REFERENCES

(1) Fraenkel-Conrat, H.; Williams, R. C. Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components. *Proc. Natl. Acad. Sci. U. S. A.* **1955**, *41* (10), 690−698.

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002

(2) Bancroft, J. B.; Hiebert, E. Formation of an Infectious Nucleoprotein from Protein and Nucleic Acid Isolated from a Small Spherical Virus. *Virology* **1967**, *32* (2), 354−356.
(3) Bancroft, J. B.; Hiebert, E.; Bracker, C. E. The Effects of Various Polyanions on Shell Formation of Some Spherical Viruses. *Virology* **1969**, *39* (4), 924−930.
(4) Johnson, J. M.; Willits, D. A.; Young, M. J.; Zlotnick, A. Interaction with Capsid Protein Alters RNA Structure and the Pathway for in Vitro Assembly of Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* **2004**, *335* (2), 455−464.
(5) Cadena-Nava, R. D.; Comas-Garcia, M.; Garmann, R. F.; Rao, A. L. N.; Knobler, C. M.; Gelbart, W. M. Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High protein/RNA Mass Ratio. *J. Virol.* **2012**, *86* (6), 3318−3326.
(6) Comas-Garcia, M.; Cadena-Nava, R. D.; Rao, A. L. N.; Knobler, C. M.; Gelbart, W. M. In Vitro Quantification of the Relative Packaging Efficiencies of Single-Stranded RNA Molecules by Viral Capsid Protein. *J. Virol.* **2012**, *86* (22), 12271−12282.
(7) Garmann, R. F.; Comas-Garcia, M.; Gopal, A.; Knobler, C. M.; Gelbart, W. M. The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. *J. Mol. Biol.* **2014**, *426* (5), 1050−1060.
(8) Garmann, R. F.; Comas-Garcia, M.; Koay, M. S. T.; Cornelissen, J. J. L. M.; Knobler, C. M.; Gelbart, W. M. Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus. *J. Virol.* **2014**, *88* (18), 10472−10479.
(9) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Packaging Signals in Two Single-Stranded RNA Viruses Imply a Conserved Assembly Mechanism and Geometry of the Packaged Genome. *J. Mol. Biol.* **2013**, *425* (17), 3235−3249.
(10) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Building a Viral Capsid in the Presence of Genomic RNA. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2013**, *87* (2), 1−12.
(11) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Solving a Levinthal's Paradox for Virus Assembly Identifies a Unique Antiviral Strategy. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (14), 5361−5366.
(12) Gan, H. H.; Fera, D.; Zorn, J.; Shiffeldrim, N.; Tang, M.; Laserson, U.; Kim, N.; Schlick, T. RAG: RNA-As-Graphs Database - Concepts, Analysis, and Features. *Bioinformatics* **2004**, *20* (8), 1285−1291.
(13) Yoffe, A. M.; Prinsen, P.; Gopal, A.; Knobler, C. M.; Gelbart, W. M.; Ben-Shaul, A. Predicting the Sizes of Large RNA Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (42), 16153−16158.
(14) Fang, L. T.; Gelbart, W. M.; Ben-Shaul, A. The Size of RNA as an Ideal Branched Polymer. *J. Chem. Phys.* **2011**, *135* (15), 155105.
(15) Fang, L. T.; Yoffe, A. M.; Gelbart, W. M.; Ben-Shaul, A. A Sequential Folding Model Predicts Length-Independent Secondary Structure Properties of Long ssRNA. *J. Phys. Chem. B* **2011**, *115* (12), 3193−3199.
(16) Yoffe, A. M.; Prinsen, P.; Gelbart, W. M.; Ben-Shaul, A. The Ends of a Large RNA Molecule Are Necessarily Close. *Nucleic Acids Res.* **2011**, *39* (1), 292−299.
(17) Gopal, A.; Egecioglu, D. E.; Yoffe, A. M.; Ben-Shaul, A.; Rao, A. L. N.; Knobler, C. M.; Gelbart, W. M. Viral RNAs Are Unusually Compact. *PLoS One* **2014**, *9* (9), e105875.
(18) Adolph, K. W.; Butler, P. J. Studies on the Ssembly of a Spherical Plant Virus III Reassembly of Infectious Virus under Mild Conditions. *J. Mol. Biol.* **1977**, *109*, 345−357.
(19) Zeng, Y.; Larson, S. B.; Heitsch, C. E.; McPherson, A.; Harvey, S. C. A Model for the Structure of Satellite Tobacco Mosaic Virus. *J. Struct. Biol.* **2012**, *180* (1), 110−116.
(20) Tzlil, S.; Ben-Shaul, A. Flexible Charged Macromolecules on Mixed Fluid Lipid Membranes: Theory and Monte Carlo Simulations. *Biophys. J.* **2005**, *89* (5), 2972−2987.
(21) Tzlil, S.; Murray, D.; Ben-Shaul, A. The "Electrostatic-Switch" Mechanism: Monte Carlo Study of MARCKS-Membrane Interaction. *Biophys. J.* **2008**, *95* (4), 1745−1757.
(22) Rosenbluth, M. N.; Rosenbluth, A. W. Monte Carlo Simlations of the Average Extension of Molecular Chains. *J. Chem. Phys.* **1955**, *23*, 356−359.
(23) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: New York, 1996.
(24) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21* (6), 1087−1092.
(25) Johnson, J. M.; Tang, J.; Nyame, Y.; Willits, D.; Young, M. J.; Zlotnick, A. Regulating Self-Assembly of Spherical Oligomers. *Nano Lett.* **2005**, *5* (4), 765−770.
(26) Zlotnick, A. Theoretical Aspects of Virus Capsid Assembly. *J. Mol. Recognit.* **2005**, *18* (6), 479−490.
(27) Zandi, R.; van der Schoot, P. Size Regulation of Ss-RNA Viruses. *Biophys. J.* **2009**, *96* (1), 9−20.
(28) Voorhees, P. W. Ostwald Ripening of Two-Phase Mixtures. *Annu. Rev. Mater. Sci.* **1992**, *22*, 197−215.
(29) Zinke-Allmang, M.; Feldman, L. C.; Grabow, M. H. Clustering on Surfaces. *Surf. Sci. Rep.* **1992**, *16* (8), 377−463.
(30) Prüfer, H. Neuer Beweis Eines Satzes Über Permutationen. *Arch. Math. Phys.* **1918**, *27*, 742−744. See also https://en.wikipedia.org/wiki/Pr%C3%BCfer_sequence
(31) Comas-Garcia, M.; Garmann, R. F.; Singaram, S. W.; Ben-Shaul, A.; Knobler, C. M.; Gelbart, W. M. Characterization of Viral Capsid Protein Self-Assembly around Short Single-Stranded RNA. *J. Phys. Chem. B* **2014**, *118* (27), 7510−7519.
(32) Perlmutter, J. D.; Hagan, M. F. The Role of Packaging Sites in Efficient and Specific Virus Assembly. *J. Mol. Biol.* **2015**, *427*, 2451.
(33) Perlmutter, J. D.; Qiao, C.; Hagan, M. F. Viral Genome Structures Are Optimal for Capsid Assembly. *eLife* **2013**, *2*, 00632.
(34) Erdemci-Tandogan, G.; Wagner, J.; Van Der Schoot, P.; Podgornik, R.; Zandi, R. RNA Topology Remolds Electrostatic Stabilization of Viruses. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2014**, *89* (3), 032707.
(35) Wu, B.; Grigull, J.; Ore, M. O.; Morin, S.; White, K. A. Global Organization of a Positive-Strand RNA Virus Genome. *PLoS Pathog.* **2013**, *9* (5), e1003363.
(36) Athavale, S. S.; Gossett, J. J.; Bowman, J. C.; Hud, N. V.; Williams, L. D.; Harvey, S. C. In Vitro Secondary Structure of the Genomic RNA of Satellite Tobacco Mosaic Virus. *PLoS One* **2013**, *8* (1), e54384.
(37) Archer, E. J.; Simpson, M. a.; Watts, N. J.; O'Kane, R.; Wang, B.; Erie, D. a.; McPherson, A.; Weeks, K. M. Long-Range Architecture in a Viral RNA Genome. *Biochemistry* **2013**, *52* (18), 3182−3190.
(38) Garmann, R. F.; Gopal, A.; Athavale, S. S.; Knobler, C. M.; Gelbart, W. M.; Harvey, S. C. Visualizing the Global Secondary Structure of a Viral RNA Genome with Cryo-Electron Microscopy. *RNA* **2015**, *21* (5), 877−886.

DOI: 10.1021/acs.jpcb.5b06445
J. Phys. Chem. B 2015, 119, 13991−14002