

RESEARCH ARTICLE

Packaging contests between viral RNA molecules and kinetic selectivity

Inbal Mizrahi¹, Robijn Bruinsma^{1,2*}, Joseph Rudnick¹**1** Department of Physics and Astronomy, University of California, Los Angeles, California, United States of America, **2** Department of Chemistry and Biochemistry, University of California, Los Angeles, California, United States of America* bruinsma@physics.ucla.edu

OPEN ACCESS

Citation: Mizrahi I, Bruinsma R, Rudnick J (2022) Packaging contests between viral RNA molecules and kinetic selectivity. PLoS Comput Biol 18(4): e1009913. <https://doi.org/10.1371/journal.pcbi.1009913>**Editor:** Samuel Coulbourn Flores, Stockholm University, Sweden, SWEDEN**Received:** August 15, 2021**Accepted:** February 9, 2022**Published:** April 1, 2022**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1009913>**Copyright:** © 2022 Mizrahi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files.**Funding:** This study was funded by National Science Foundation, Directorate for Mathematical

Abstract

The paper presents a statistical-mechanics model for the kinetic selection of viral RNA molecules by packaging signals during the nucleation stage of the assembly of small RNA viruses. The effects of the RNA secondary structure and folding geometry of the packaging signals on the assembly activation energy barrier are encoded by a pair of characteristics: the *wrapping number* and the *maximum ladder distance*. Kinetic selection is found to be optimal when assembly takes place under conditions of *supersaturation* and also when the concentration ratio of capsid protein and viral RNA concentrations equals the stoichiometric ratio of assembled viral particles. As a function of the height of the activation energy barrier, there is a form of order-disorder transition such that for sufficiently low activation energy barriers, kinetic selectivity is erased by entropic effects associated with the number of assembly pathways.

Author summary

During the assembly of a viral particle, a limited number of viral genomic RNA molecules must compete for packaging with a large number of closely similar host messenger RNA molecules. All-atom simulations of this competition process are impractical. The paper presents a tractable mathematical model for the selection process as a *non-equilibrium* phenomenon.

Introduction

When the molecular components of a single-stranded (ss) RNA virus assemble and form virions in the cytoplasm of an infected cell, genomic viral RNA molecules (gRNA) compete for packaging with a large pool of—quite similar—host messenger RNA (mRNA) molecules for packaging by the viral capsid proteins [1]. For example, for the case of influenza the number of gRNA molecules inside an infected cell is less than 10^4 [2] while the total number of host mRNA molecules is in the range of 3.6×10^5 .

and Physical Sciences, CMMT Grant No.1836404. https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505357 (to R.B.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Viral RNA selection relies on so-called *packaging signals* [3–9]. These are short RNA stem loop motifs that are a part of the secondary structure of gRNA molecules. Importantly, *any* ssRNA molecule with the molecular weight of a gRNA molecule has numerous hairpins, some of which may be similar or identical to one of the packaging signals of a virus. In order to avoid the packaging of mRNA material *individual* packaging signals should not trigger virion assembly. Viral gRNA molecules typically have a coordinated pattern of packaging signals that collectively direct the assembly, sometimes called a “ ψ sequence”. These virus-specific interactions between packaging signals and capsid proteins operate together with a generic, non-specific electrostatic affinity between the negatively charged RNA nucleotides and positively charged residues of the capsid proteins [10–13].

It has been demonstrated that the spontaneous self-assembly of empty icosahedral capsids is initiated by the formation of a *nucleation complex* composed of a limited number of capsid proteins [14–16]. This nucleation complex can be compared to the *critical nucleus* of the kinetic theory of nucleation and growth [17, 18]. The energetically uphill formation of the nucleation complex is followed by the energetically down-hill growth (or “elongation”) process that ends in the formation of a closed capsid. The self-assembly of empty capsids does not (and should not) take place under physiological conditions. Under those conditions electrostatic repulsion between the capsid proteins is just able to overcome the hydrophobic affinity between capsid proteins. The physical aspects of RNA packaging have been extensively studied experimentally and theoretically [10, 11, 19–35] as well as by numerical modeling [13, 36–38]. For reviews see refs. [18, 39, 40]. Theoretical models have generally focused on the minimization of the free energy of a fully assembled viral particle. This produced global measures for the “packaging fitness” of an RNA molecule in terms of its length, the degree of branching and compactness, and the effects of electrostatics and osmotic pressure.

The pioneering work by Aaron Klug on TMV [41] showed that the action of gRNA on assembly is two-fold. On the one hand, negative charges of the RNA molecules neutralize—on a non-specific basis—positive capsid protein charges. This shifts the overall equilibrium free-energy balance from a dispersed state towards aggregation. On the other hand, the specific packaging signals on gRNA act as *catalysts* that lower the activation energy barrier of the nucleation complex. In the view of Klug, the packaging signals affect the assembly *kinetics* while in a thermodynamic view the role of the packaging signals would be to further tilt the free energy balance in favor of packaging. The most well-studied case of RNA selectivity is probably that of the HIV-1 retrovirus (see ref. [42] and references therein). RNA selectivity depends on the cooperative action of a cluster of packaging signals located at the 5' end of the gRNA molecule, the ψ sequence. It is about a hundred nucleotides long, which is very small compared to the total length of the HIV-1 genome (about 10^4 nucleotides). gRNA selection appears to take place very early, during the nucleation stage of the assembly process when the ψ sequence interacts with only a small group of capsid proteins. Changing the RNA sequence of other sections of the genome molecules does not appear to affect the selectivity. *The gRNA molecules appear to have no thermodynamic advantage over non-viral RNA molecules of the same length* [43]. This indicates that the origin of the very efficient gRNA selection mechanism of HIV-1 must be sought in the kinetics of the assembly process. Finally, recent progress in the asymmetric image reconstruction of certain small RNA viruses [44] indicate that also in those cases the RNA selection process takes place early in the assembly process. Asymmetric reconstruction of the MS2 phage virus [45] shows that RNA packaging signals associate *reproducibly* with a specific section of the interior of the capsid. The authors proposed a model for viral assembly in which a spatial distribution of packaging signals functions as a virus-specific “map” for the initial nucleation stage of the assembly while the subsequent elongation step of the assembly is driven more by non-specific interactions. This scenario appears similar to that

of the TMV and HIV-1 viruses. There certainly are also counter examples where free energy minimization accounts for selection. For example, the asymmetric reconstruction of the CCMV and BMV plant viruses produced only a very small amount of reproducible RNA-protein association [46]. Interestingly, this group of viruses is much less selective than MS2. CCMV capsid proteins appear to select BMV genome molecules over CCMV genome molecules while they can package a wide variety of non-viral ss RNA molecules and even non-RNA polyelectrolytes [47, 48]. Apparently, the amount of CCMV gRNA molecules produced inside an infected cell is sufficiently large so there is no need for very precise assembly selectivity.

In this paper we propose a simple statistical-physics model to study the physics of selective nucleation by a group of packaging signals that encode the assembly of a small ssRNA virus. By construction, the model focuses exclusively on kinetic selection. In the conclusion we will return to the relation between the thermodynamic and kinetic modes of selection.

Model and methods

Spanning tree model

The starting state of the system is assumed to be a solution containing a certain concentration of condensed, folded viral RNA molecules and of pentameric capsid proteins. The folded molecules have the same interior structure and dimensions and differ only in terms of the ψ section of contiguous packaging signals distributed over the surface of the condensed RNA molecule. The capsid of the virus is assumed to be composed of twelve protein pentamers assembled into a dodecahedron such that double-stranded (ds) RNA sequences line the edges of the pentamers. This is inspired by the family of the *Nodaviridae* in which part of the ssRNA genome are ds sequences forming a dodecahedral cage [49]. The secondary structure of the ψ section is represented as a tree of nineteen links and twenty nodes connecting the vertices of the dodecahedron. The pentameric capsid itself is the well-studied Zlotnick model system for empty capsids [50–53]). The geometry of the ψ section is assumed to be adapted to the dodecahedral capsid so that the twenty nodes of the secondary structure match up with the twenty vertices of the dodecahedron. Despite these constraints there still are tens of thousands of secondary structures that satisfy these constraints. In the mathematical literature, these structures are known as the *spanning tree graphs* of a dodecahedron [54]. The number of nodes of a spanning tree is twenty because a spanning tree must visit all the vertices, of which there are twenty. The number of links is nineteen because in any connected tree graph the number of links is one less than the number of nodes. A spanning tree leaves eleven edges of the dodecahedron uncovered. We will assume that these remaining edges have only a generic affinity for the capsid proteins. Fig 1 shows an example of a spanning tree of the dodecahedron.

Initially, all pentamers are in solution at a certain total concentration c_0 . The pentamers are assumed to have a generic affinity (electrostatic in actuality) for all edges of the dodecahedron plus a specific affinity for those edges that are covered by links of the spanning tree (acting as packaging signals). The specific affinity is maximized by placing the pentamer on a location such that four of its edges can associate with a link of the chain. For the tree molecule shown in Fig 1, a total of six pentamers can be placed on such maximum wrapping locations. We will say that the *wrapping number* of this tree structure is $N_p = 6$. The wrapping number is a characteristic of the folding geometry of the RNA molecule.

The very simplest spanning tree is a linear chain composed of nineteen links. Fig 2 shows how a linear chain can be distributed over a dodecahedron, while visiting all vertices. As shown in Fig 2, only two pentamers can be placed on locations such that four of its edges associate with a link of the chain. There are 1620 different configurations for a linear chain of twenty nodes to be distributed over of a dodecahedron such that the nodes coincide with the vertices

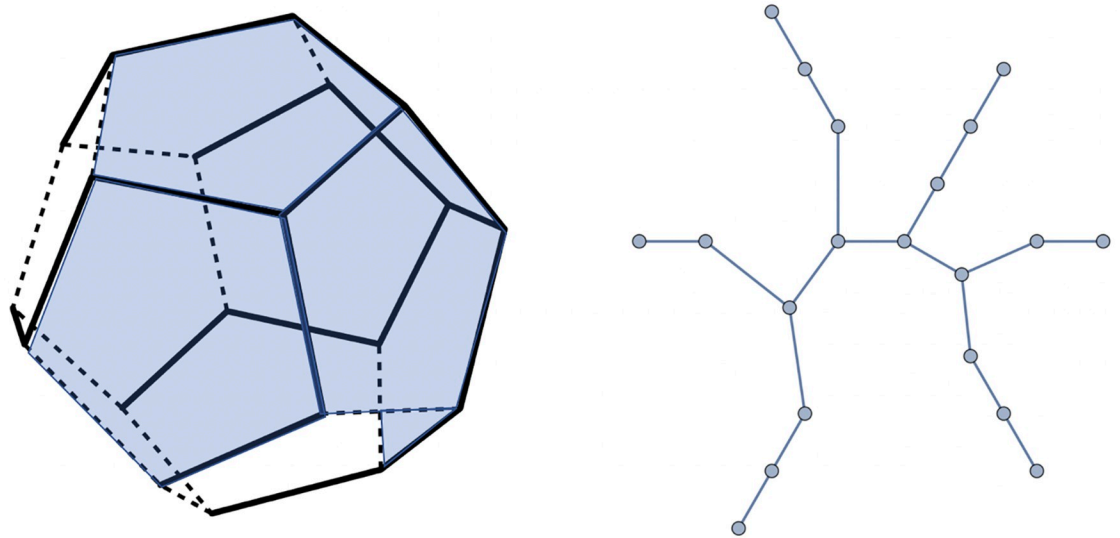


Fig 1. Left: Branched spanning tree connecting the vertices of a dodecahedron. Six pentamers can be placed on the dodecahedron such that their edges make the maximum of four contacts with links of the spanning tree. Right: Planar representation of the spanning tree.

<https://doi.org/10.1371/journal.pcbi.1009913.g001>

known in the mathematical literature as *Hamiltonian paths*. Hamiltonian paths have been used before to classify viral RNA configurations [8, 55, 56]). Hamiltonian paths of the dodecahedron can have wrapping numbers two, three, or four. There are important differences between the linear and the branched cases. In the linear case, pentamers have embeddings with two, three or four edges occupied while for the branched case, there are embeddings with zero, one, two, three or four edges occupied.

Next, the edges of a pentamer will be assumed to have affinity for the edges of other pentamers. It is this affinity that drives the assembly of empty capsids. The wrapping number does

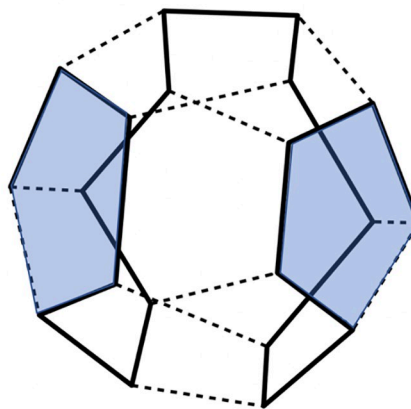


Fig 2. Example of a spanning tree in the form of a Hamiltonian path of a dodecahedron. In blue are shown two pentamers that can be placed on the dodecahedron such that their edges make the maximum of four contacts with links of the spanning tree.

<https://doi.org/10.1371/journal.pcbi.1009913.g002>

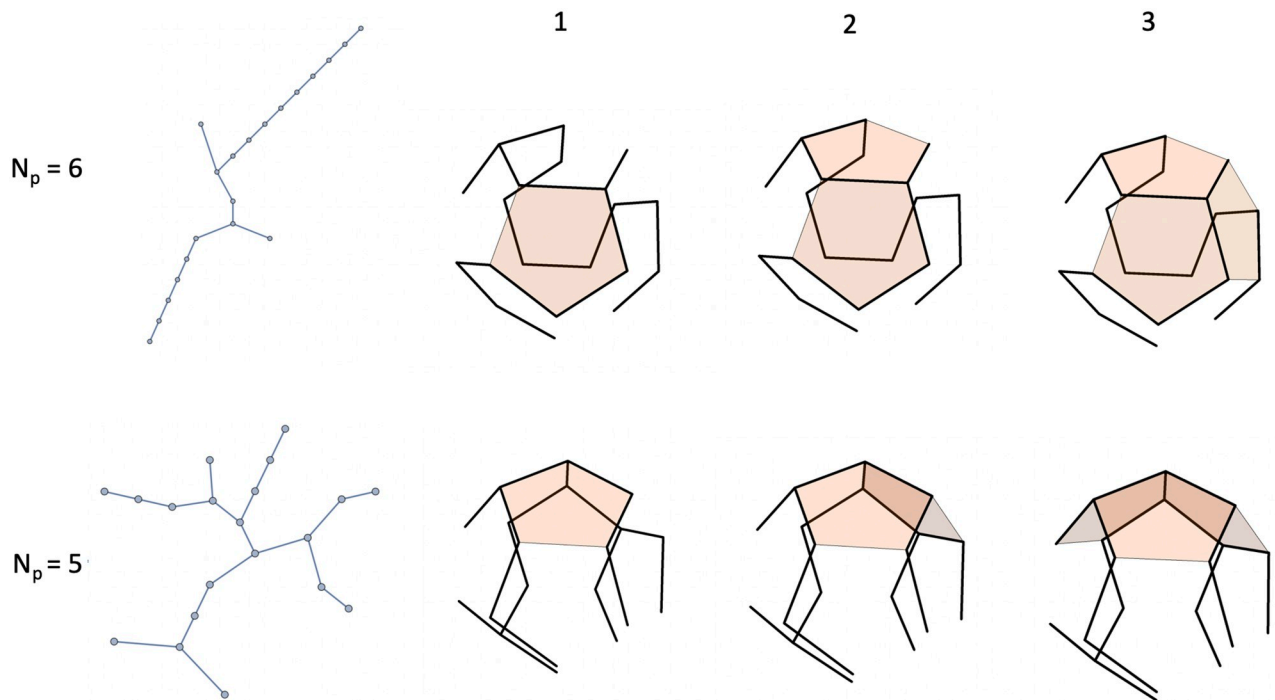


Fig 3. The first three steps of the minimum energy assembly pathways of two molecules, with wrapping numbers $N_p = 6$ (top) and $N_p = 5$ (bottom) respectively, for the case that the affinity of pentamer edges for each other exceeds the specific affinity for the spanning tree. The more compact $N_p = 5$ tree allows three pentamers on sites with four links with each pentamer making two edge-to-edge contacts. For the less compact $N_p = 6$ tree, the third pentamer only makes three contacts with a link.

<https://doi.org/10.1371/journal.pcbi.1009913.g003>

not measure how many attractive contacts a newly added pentamer can make with pentamers that were placed earlier on the dodecahedron. The more *compact* a spanning tree, the larger the probability that two maximally wrapped pentamers also are able to share an edge. Fig 3 illustrates the important role of compactness of the spanning tree. The figure shows the first three steps of the minimum energy assembly pathways of two spanning trees, one with a wrapping numbers $N_p = 5$ and the other with $N_p = 6$. It is assumed that the edge-edge affinity exceeds the edge-link affinity. In both cases, the first two pentamers can be placed on adjacent sites with maximum wrapping so they have one shared edge. The assembly energy is the same at this point. A difference appears when the third pentamer is placed on the assembly. For the more compact $N_p = 5$ molecule, the third pentamer can be placed on a maximum wrapping site where it has two shared edges with the two pentamers already present but this is not possible for the less compact $N_p = 6$ tree. It follows that the minimum assembly energy of a three pentamer cluster for the $N_p = 5$ molecule is lower than that of the $N_p = 6$ molecule. The wrapping number is thus, by itself, insufficient as an index that can predict which spanning trees favor assembly nucleation.

The Maximum Ladder Distance (MLD) has been used to characterize the degree of compactness of the secondary structure of complete gRNA molecules and as a measure of the size of an RNA molecule in solution [57, 58]. In the mathematical literature, the ladder distance (or LD) between two nodes of a tree graph is defined as the number of links of the graph along a minimum length path separating the two nodes. The MLD of a tree graph is the largest LD of the graph. In the language of graph theory, the LD between two nodes of a tree graph is known as the “distance” between the two nodes and the MLD as the “diameter” of a tree graph [59].

Below we apply the MLD concept only to the ψ section of the complete molecules, not to the RNA molecule as a whole. The MLD of the branched spanning tree that was just discussed is nine while it is nineteen for the linear tree. Unlike the wrapping number, the MLD is a characteristic that is determined entirely by the topology of the planar graph of the secondary structure of the ψ sequence. Unlike the wrapping number, it remains the same for different folding geometries. We will see that the combination of the wrapping and MLD numbers together forms a satisfactory index for the effectiveness of a spanning tree as a nucleation catalyst.

Minimum-energy assemblies

A spanning tree/pentamer structure with n pentamers is assigned an assembly energy

$$\Delta E(n) = E_0(n_1\epsilon_1 + n_2\epsilon_2 - n_3 - \mu_0 n)$$

with respect to an RNA molecule without pentamers. Here, n_1 is the number of links of the spanning tree that lie along a pentamer edge that is not shared with another pentamer, n_2 is the number of spanning tree links that lie along a pentamer edge that is shared with another pentamer while n_3 is the number of edges shared between two pentamers that are not associated with a link of the spanning tree. The energy scale E_0 is the binding energy between two pentamer edges in the absence of specific RNA-pentamer affinity. Energies will be expressed in units of E_0 . The physical meaning of the dimensionless parameter $-\epsilon_1$ is that of the ratio of the affinity of a spanning tree link for a pentamer edge over E_0 . Interactions between edges and spanning tree links will be assumed to be additive. In that case the dimensionless ϵ_2 parameter equals $\epsilon_2 = -1 + 2\epsilon_1$ since the RNA link interacts with two pentamer edges. Finally, μ_0 is the chemical potential of pentamers in solution at a certain reference concentration. The reference chemical potential includes the non-specific affinity of a pentamer for the RNA condensate. The reference chemical potential will be chosen so that $\Delta E(12)$ is close to zero, so when the chemical potential in solution at the reference concentration is the same as that of a pentamer that is part of an assembled capsid. This is the case if $\Delta E(12) = 19\epsilon_2 - 11 - 12\mu_0$ is zero. Note that $\Delta E(12)$ is the same for all spanning trees so different spanning trees have the same assembly energy (we will see later that the assembly *free* energy is not the same for all trees).

Two examples of minimum-energy assembly profiles near assembly equilibrium with $c_0 = 1$ are shown in Fig 4. The left figure shows the case of an $N_p = 8$, $MLD = 9$ spanning tree. Recall that such a spanning tree is maximally adjusted for pentamer binding. The right figure shows the case of an $N_p = 2$, $MLD = 19$ spanning tree, which has minimal adjustment for pentamers. The activation energy is about two E_0 higher in the second case. The assembly energy profiles of spanning trees with the same N_p and MLD are nearly always the same.

These assembly energy profiles are consistent with a nucleation-and-growth scenario close to the equilibrium assembly threshold. As expected, the activation energy barrier of the $N_p = 8$, $MLD = 9$ spanning tree (about $3.0E_0$) is lower than that of the $N_p = 2$, $MLD = 19$ spanning tree (about $5.0E_0$). The long straight section of the $N_p = 8$, $MLD = 9$ energy profile can be understood by noting that when a pentamer is added to one of the eight maximum wrapping sites then that lowers the energy by $4\epsilon_1$ in units of E_0 . If additional pentamers always make two new contacts with pentamers that are already present—as is indeed the case here—then each added pentamer lowers the energy further by an amount of $2E_0$ minus the chemical potential μ_0 . In this particular case, these two terms cancel so the assembly energy barrier has a wide and flat top. Starting from a linear chain with $MLD = 19$ and $N_p = 2$ and then stepwise decreasing the MLD and increasing the N_p one finds that the assembly energy activation barrier nearly always systematically decreases. Assembly on a spanning tree with the minimum MLD and the

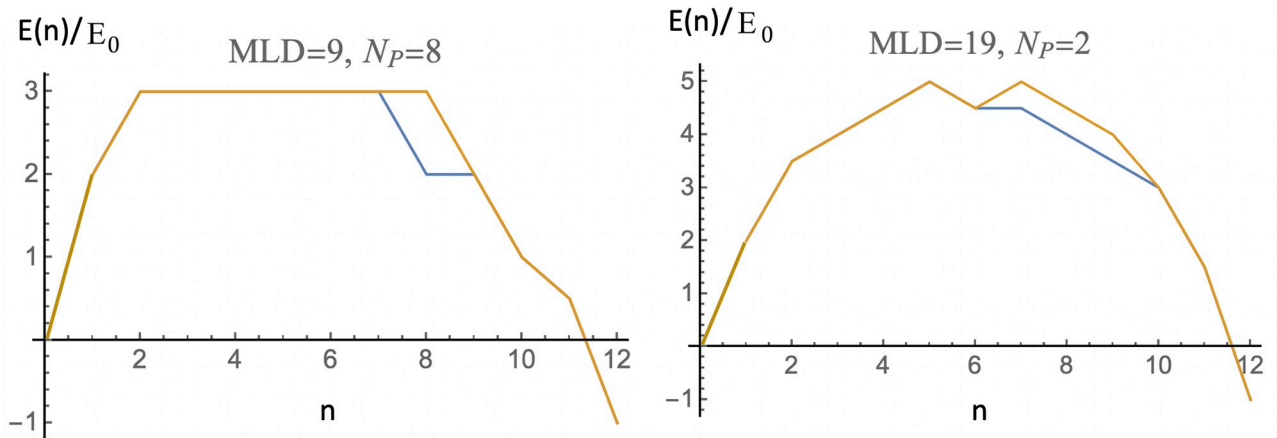


Fig 4. Minimum-energy assembly energy profiles for $N_p = 8$, $MLD = 9$ spanning trees (left) and for an $N_p = 2$, $MLD = 19$ spanning tree (right). Energy parameters are $\epsilon_1 = -0.5$ and $\mu_0 = -4.0$ (close to assembly equilibrium where $\mu_0 \simeq -4.083$). Energies are expressed in units of the overall scale E_0 . The assembly energy profiles of spanning trees with the same N_p and MLD are nearly always the same. The pathways for the small number of exceptions are shown in the figure.

<https://doi.org/10.1371/journal.pcbi.1009913.g004>

maximal wrapping number N_p has in general the lowest possible assembly activation energy barrier.

An illustration of the assembly sequence of an $N_p = 8$, $MLD = 9$ spanning tree spanning tree is shown in Fig 5 for the case that $-\epsilon_1$ is less than 0.5. The first eight pentamers all can be placed on sites that maximize the number of spanning tree link contacts (four) as well as the number of attractive pentamer-pentamer contacts. The second pentamer creates one new pentamer-pentamer contact while the next three pentamers create two new pentamer-pentamer contacts. All pentamer assembly intermediates are compact structures. In summary, the combination of the wrapping number and the MLD appears to be a good, though not perfect, code for the height of the assembly energy activation barrier.

Next, we explored the configuration space of assembly intermediates. Fig 6 shows graphs of all distinct assembly intermediates for molecule (1) and (2) as well as the minimum energy assembly pathways that link them. Each node of the network stands here for a physically distinct assembly intermediate (assemblies that are related by a symmetry operation of the dodecahedron are treated here as the same). Nodes are assigned “coordinates” (n, i) with $n = 0, 1, \dots, 12$ the number of pentamers of the intermediate and with $i = 1, 2, \dots, m_n$ an index ranging over the distinct n -pentamer states where m_n is the *multiplicity* of the n -pentamer state (e.g., $m_5 = 4$ for the $N_p = 8$, $MLD = 9$ spanning tree). A black line linking two dots indicates that the two states can be interconverted by addition or removal of a pentamer. Assembly of viral particles can be viewed as a net “current” flow from the $n = 0$ source state to the $n = 12$ final state along all possible paths across the network linking the initial state to the final state. Under conditions of thermodynamic equilibrium, the current across each individual link should be zero according to the principle of detailed balance. Note that the compact, branched spanning tree molecule (1) (left) has far fewer assembly intermediates and assembly pathways than the linear structure molecule (2) (right).

An important simplification ensues when we ignore the very small number of assembly energy profiles that do not conform to the quasi-universal profile for given N_p and MLD discussed above. If we do that then *different nodes of the network with the same n all have the same assembly energy $\Delta E(n)$* . This simplification allows us assign to each node a Boltzmann

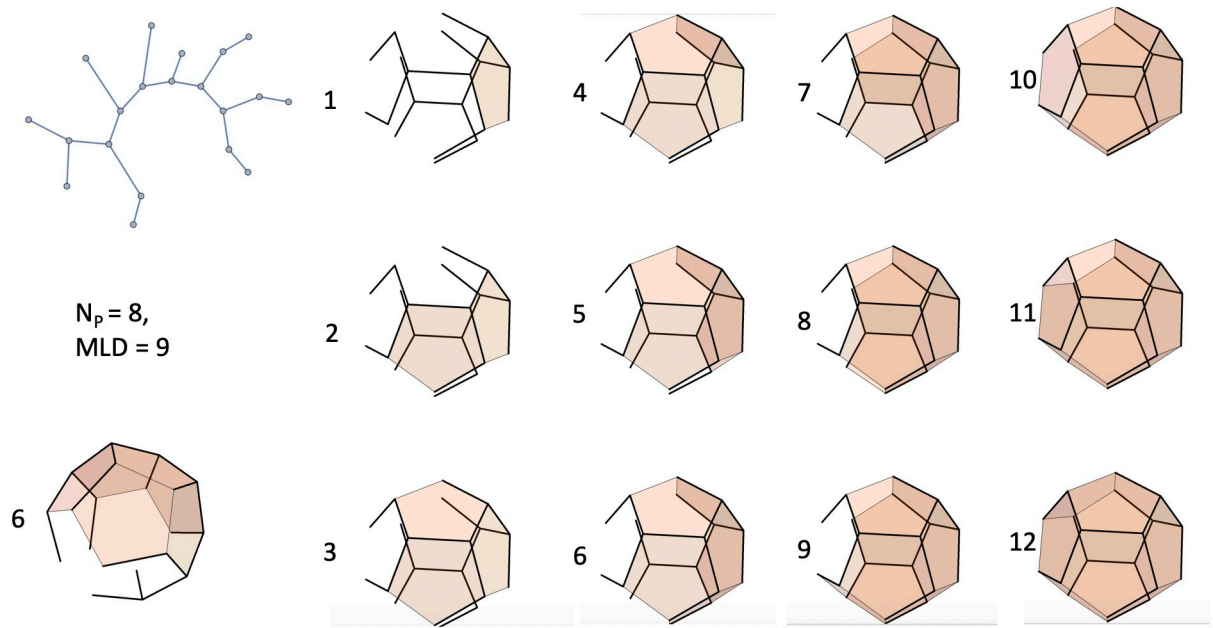


Fig 5. A spanning tree with a maximum wrapping number of eight and the minimum MLD of nine. The first five pentamers can be placed on sites that maximize both the number of available pentamer-pentamer contacts and pentamer-spanning tree link contacts (four). The sixth pentamer, shown separately with a different perspective, makes three pentamer-pentamer contacts but can only make two spanning tree link contacts.

<https://doi.org/10.1371/journal.pcbi.1009913.g005>

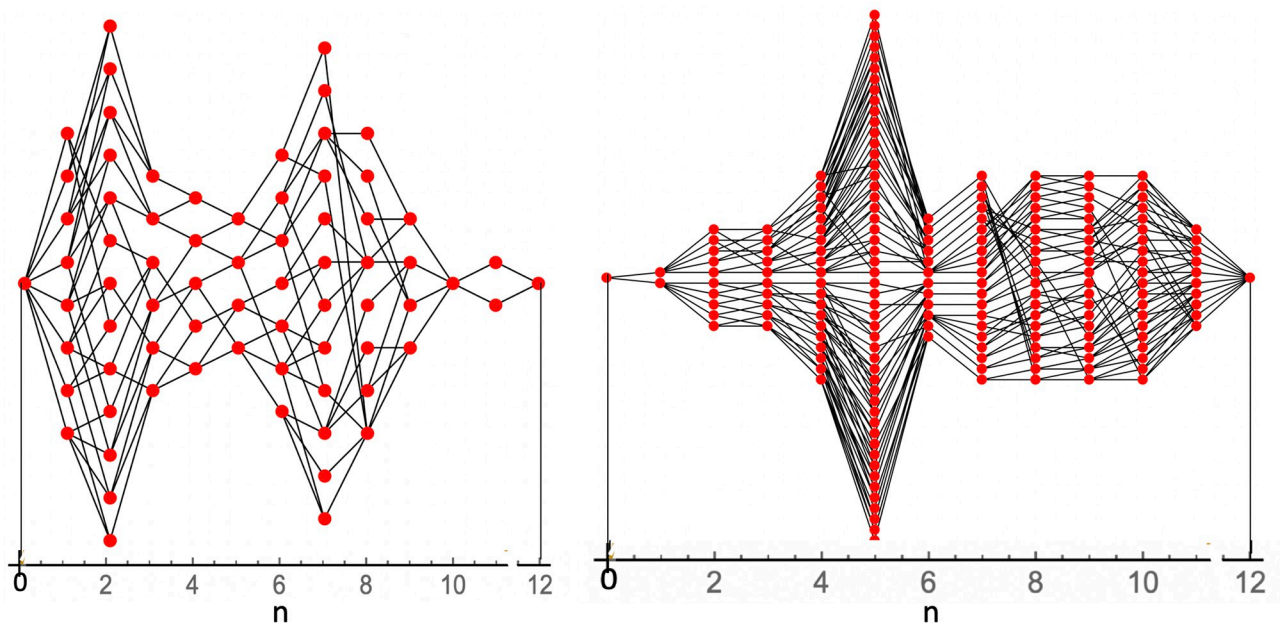


Fig 6. Examples of minimum energy assembly paths of an $N_p = 8$, $MLD = 9$ spanning tree (left) and an $N_p = 2$, $MLD = 19$ spanning tree (right). A node (indicated by a dot) indicates a physically distinct intermediate structure with, from left to right, $n = 0, 1, \dots, 12$ pentamers. The number m_n of vertical dots for given n is the *multiplicity*, i.e., the number of distinct n -pentamer intermediates. A link connecting two nodes indicates that the two states are related by addition or removal of a pentamer. Every possible path from $n = 0$ to $n = 12$, including back steps, represents a possible minimum energy assembly pathway. During assembly there is a net current from the $n = 0$ initial state to the $n = 12$ final state while in thermal equilibrium the net current across every link is zero.

<https://doi.org/10.1371/journal.pcbi.1009913.g006>

equilibrium probability:

$$P_{eq}(n) \propto \exp - \beta \Delta E(n) + n \ln(c_f) \tag{1}$$

which depends on the concentration c_f of pentamers free in solution (expressed in units of the reference concentration). The proportionality factor in the expression is determined by the normalization condition $\sum_{n=1}^{12} m_n P_{eq}(n) = 1$. We will make this simplification in the following sections.

Master equation

In this section we define the Master Equation that governs the kinetics. We will use the coordinate system for the network graphs defined below Fig 6. The network geometry of a particular spanning tree is specified in the form of an adjacency matrix A_n^{ij} that equals one if a link connects node n, i to node $n + 1, j$ while it equals zero if there is no link. Each node n, i of the graphs has a time-dependent occupation probability $P_{i,n}(t)$. The kinetics is expressed as a set of coupled rate equations for the $P_n^i(t)$ and is assumed to be a Markov process with probabilities evolving in time according to the Master Equation [60]:

$$\begin{aligned} \frac{dP_{i,n}(t)}{dt} = & \sum_j \{ A_{n-1}^{ji} W_{n-1,n} P_{j,n-1}(t) + A_n^{ij} W_{n+1,n} P_{j,n+1}(t) \} \\ & - P_{i,n}(t) \sum_j \{ A_{n-1}^{ji} W_{n,n-1} + A_n^{ij} W_{n,n+1} \} \end{aligned} \tag{2}$$

Here, $W_{n,n+1}$ is the on-rate for the transition of an assembly of n pentamers to one with size $n + 1$ by the addition of a pentamer while $W_{n,n-1}$ is the off-rate at which a pentamer is removed from an assembly of size n . We assume a simplified diffusion-limited chemical kinetics (see [61] and Supplementary Information S1 Text) in which the addition or removal of a pentamer to an assembly of size n is treated as a bimolecular reaction. The resulting on-rate has the form of a kinetic Monte-Carlo algorithm:

$$W_{n,n+1} = \lambda c_f \begin{cases} e^{-\Delta \Delta E_{n,n+1}} & \text{if } \Delta E(n + 1) > \Delta E(n) \\ 1 & \text{if } \Delta E(n + 1) < \Delta E(n) \end{cases} \tag{3}$$

with c_f again the concentration of free pentamers, $\Delta \Delta E_{n,n+1} = \Delta E(n + 1) - \Delta E(n)$ the energy cost of adding a pentamer, and λ a base rate that depends on molecular quantities such as diffusion coefficients and reaction radii but not on the concentrations. The inverse of λ is the fundamental time-scale of the kinetics. In the following, time will be expressed in units of $1/\lambda$. If $\Delta \Delta E_{n,n+1}$ is negative then the on-rate is equal to this base rate. If $\Delta \Delta E_{n,n+1}$ is positive then the base rate is reduced by the Arrhenius factor $e^{-\Delta \Delta E_{n,n+1}}$. The off-rate entries $W_{n+1,n}$ are determined by the on-rates through the condition of detailed balance:

$$\frac{W_{n,n+1}}{W_{n+1,n}} = \frac{P_{eq}(n + 1)}{P_{eq}(n)} = c_f e^{\Delta \Delta E_{n,n+1}} \tag{4}$$

Below, we will limit ourselves to the case of solutions containing only two species of spanning trees, namely molecules (1) and (2), having the same solution concentrations. During assembly, the two species compete for the same concentration c_f of pentamers. The ratio of c_f

over the total pentamer concentration c_0 is determined by mass conservation:

$$c_f/c_0 = 1 - (D/24) \sum_{n=0}^{12} \left(\sum_{i=1}^{m_n^{(1)}} nP_{i,n}^{(1)}(t) + \sum_{i=1}^{m_n^{(2)}} nP_{i,n}^{(2)}(t) \right) \quad (5)$$

where superscripts denote the kind of spanning tree. Next, $D \equiv 12r_t/c_0$, with r_t the total RNA concentration, is the RNA to protein *mixing ratio*, an important thermodynamic control parameter for the assembly process. If $D = 1$ then there are exactly enough pentamers to encapsidate both types spinning trees, which corresponds to the *stoichiometric ratio*. Since the occupation probabilities that we seek to obtain enter in this relation, the rate equations form a coupled, non-linear set of two times twelve differential equations (for the case of just a single species in solution, there are twelve differential equations where one must replace $D/24$ by $D/12$).

Numerical construction and implementation of coupled master equations with competition

Coupled master equations for packaging competition between pairs of spanning trees were integrated using a Mathematica program (available on request). The program was organized as follows.

First step: Construction of spanning trees. First, we populate in all possible ways nineteen of the thirty one edges of the dodecahedron with links. Next we identify graphs that have the properties that (i) the graph is connected and that (ii) the graph has twenty vertices. All graphs having these two properties are spanning trees because (i) any connected graph with n edges and $n + 1$ vertices is a tree and (ii) they are spanning trees because a dodecahedron has twenty vertices. This method generates many duplicates so the next step is winnowing the trees down to a collection of trees that cannot be mapped into each other by any of the 120 rotations and reflections that map a dodecahedron into itself. This is a straightforward (though laborious) process that involves choosing a tree and then eliminating all other trees that can be mapped into it by a reflection or a rotation. One can speed the process up by separating trees into subsets having the same MLD, as the MLD is a topological property that is preserved under rotation and reflection.

Second step: Choice of a pair of spanning trees. Two trees are chosen from the library of all unique spanning trees of the dodecahedron. The trees are indexed by their MLD and N_p numbers.

Third step: Specification of the assembly energies. Values are assigned to the energy parameters E_0 , ϵ_1 and μ_0 , as defined above.

Fourth: Determination of the assembly network. All physically distinct minimum energy assemblies are determined forming the nodes of the assembly graph. For every node, a list is made of nodes with one more or one less pentamer from which the adjacency matrix is determined.

Fifth step: Construction of the master equation. The assembly/disassembly process is assumed to occur by the addition of a pentamer to, or removal of a pentamer from, a structure. The steps of addition or removal are controlled by two considerations. The first is the increase or decrease in the energy of the system as determined by the energy of the partially or completely assembled capsid and an assigned chemical potential. The second is the availability of pentamers, quantified by the concentration in solution of available pentamers. The free energy increment determines the relative likelihoods of addition or removal of a pentamer, via the kinetic Monte-Carlo rates that were defined above. The concentration of available

pentamers controls the rate of the assembly/disassembly process, in that disassembly takes place at a fixed rate and assembly at a rate proportional to the number of available pentamers.

Time-scales

As a preliminary, Figs 7 and 8 show the packaging kinetics of a single species of RNA molecules. The parameters are $E_0 = 4k_bT$, $\epsilon_1 = -0.5$, $c_0 = 1$ and $D = 0.5$ and $\mu_0 = -4$. Fig 7 shows the case of the class of MLD = 9 and $N_p = 8$ spanning trees and Fig 8 that of the class of MLD = 19, $N_p = 2$ spanning trees. In the late time limit, the two classes approach thermal equilibrium with roughly the same fraction of RNA molecules being packaged (about sixty six percent). This reflects the fact that all classes of molecules have—by construction—the same assembly energy (the remaining small difference is due to the fact the entropy in the assembly free energy is not the same for the two classes). The reason that in both cases a significant fraction of RNA molecules has not been packaged, despite the fact that there are twice as many pentamers as needed to package the RNA molecules, is that the chemical potential is close to assembly equilibrium so a significant number of RNA molecules remain free of pentamers. There is a large difference between the thermalization times of two classes, roughly 10^3 time units for MLD = 9 and $N_p = 8$ spanning trees and 10^5 time units for MLD = 19, $N_p = 2$ spanning trees, consistent with the fact that the assembly activation barrier is about two times E_0 larger (i.e., about $8k_bT$) for the MLD = 19, $N_p = 2$ spanning trees. These thermalization times must be compared with the assembly delay time t_d , which is defined as the time lag between the establishment of solution assembly conditions and the first appearance of assembled viral particles. Measured delay times for the assembly of empty capsids are in the range of seconds to minutes [14–16]. We obtain t_d from the intersection of the tangent to $P_{12}(t)$ at the point of maximum slope with the horizontal axis (see Fig 9).

For the case of the MLD = 9 and $N_p = 8$ class of spanning trees, this gives about 8.5 time units as shown. Other classes have comparable delay times. It follows that our time unit is roughly in the range of five seconds. The thermalization time under conditions of assembly equilibrium would then be in the range of a prohibitively long *two hundred hours* for MLD = 9, $N_p = 8$ spanning trees and two orders of magnitude longer for the MLD = 9, $N_p = 8$ spanning trees.

Packaging competition

In Figs 10, 11 and 12 we show the result of a calculation with the same total amount of RNA molecules and pentamers as before but now with half of the RNA molecules MLD = 9 and $N_p = 8$ spanning trees and the other half MLD = 19, $N_p = 2$ spanning trees.

The MLD = 9, $N_p = 8$ spanning trees dominate packaging on time scales less than about 10^7 time units while the characteristic assembly time is in the range of 10^4 time units. About eighty percent of these spanning trees are packaged. This packaged fraction then slowly decreases on time scales of the order of 10^7 – 10^8 time unit, which means that packaged MLD = 9, $N_p = 8$ spanning trees are gradually *disassembling* as a precursor of thermalization. The fraction of packaged MLD = 19, $N_p = 2$ spanning trees increases correspondingly. Apparently, pentamers that are being freed up by disassembly of MLD = 9, $N_p = 8$ spanning trees feed assembly of the MLD = 19, $N_p = 2$ spanning trees. When the system approaches thermal equilibrium, the packaging fractions of the two classes in the long-time limit are nearly the same and close to the separate equilibrium values found earlier in the absence of competition. We conclude that in a packaging competition experiment, the disassembly of particles is a crucial step for reaching thermodynamic equilibrium.

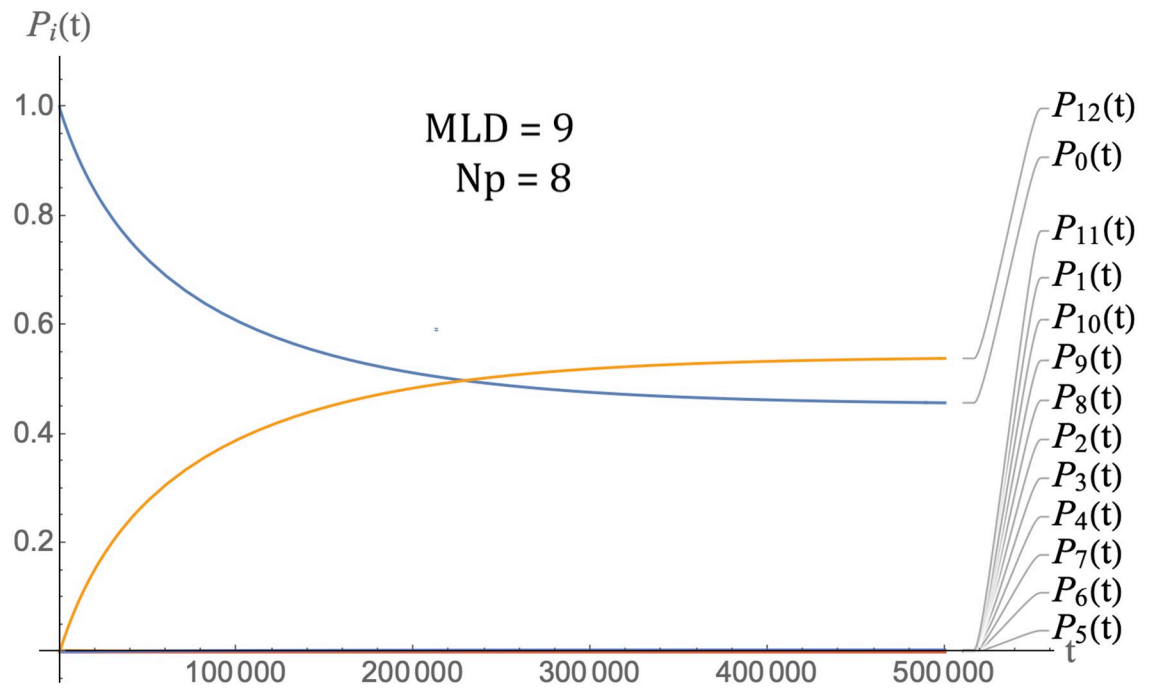


Fig 7. Packaging kinetics of MLD = 9 and $N_p = 8$ spanning trees. Parameter values are $E_0 = 4k_bT$ for the energy scale, $\epsilon_1 = -0.5$, $\epsilon_0 = 1$, $D = 0.5$ and $\mu_0 = -4$. Bottom: Packaging kinetics of MLD = 19, $N_p = 2$ spanning trees with the same parameters.

<https://doi.org/10.1371/journal.pcbi.1009913.g007>

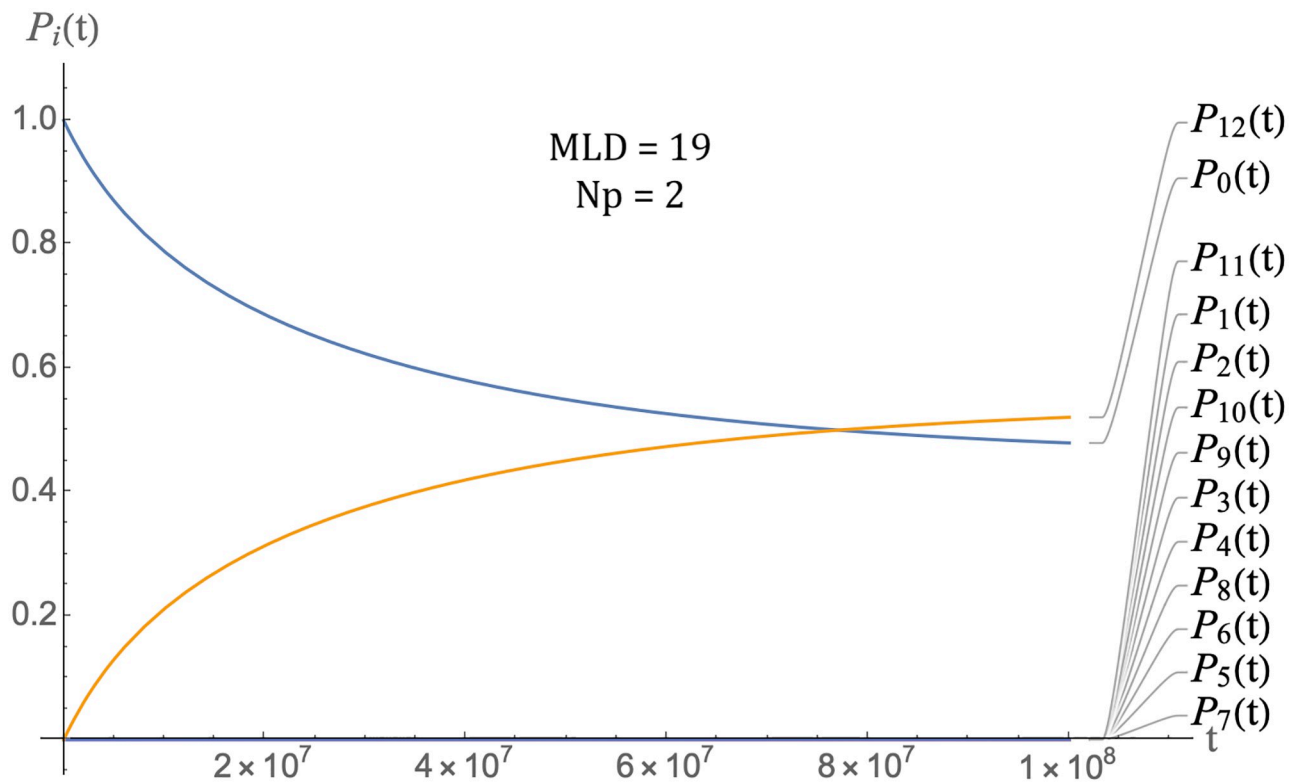


Fig 8. Packaging kinetics of MLD = 19, $N_p = 2$ spanning trees with the same parameters as those of the previous figure.

<https://doi.org/10.1371/journal.pcbi.1009913.g008>

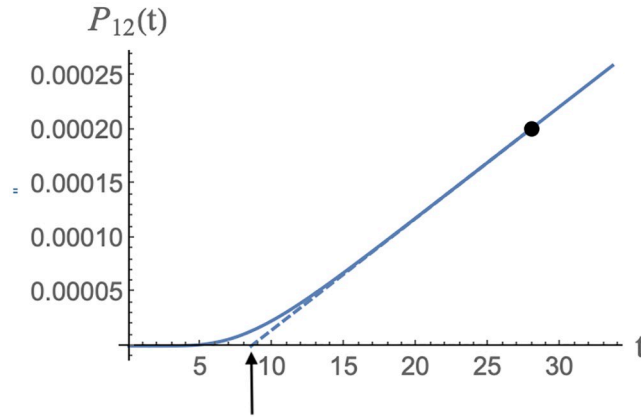


Fig 9. Definition of the delay time as the intersection of the maximum slope tangent to the assembly curve $P_{12}(t)$ with the time axis, as indicated by the arrow. $MLD = 9, N_p = 8, \epsilon = 0.5, c_0 = 1, D = 0.5$ and $\mu_0 = -4$.

<https://doi.org/10.1371/journal.pcbi.1009913.g009>

Order-disorder transitions

We now can explore how this kinetic form of RNA selection is influenced by changes in the control parameter. One reason that is necessary for us to do that is that the assembly time-scale of the $MLD = 9, N_p = 8$ spanning trees was in the range of 10^4 time units. This is much too long, of the order of five days if one uses the earlier estimate that a time unit is about five seconds. Since assembly time kinetics depends exponentially on E_0 , the assembly time can be reduced by reducing the energy scale E_0 but what happens with the selectivity if one reduces

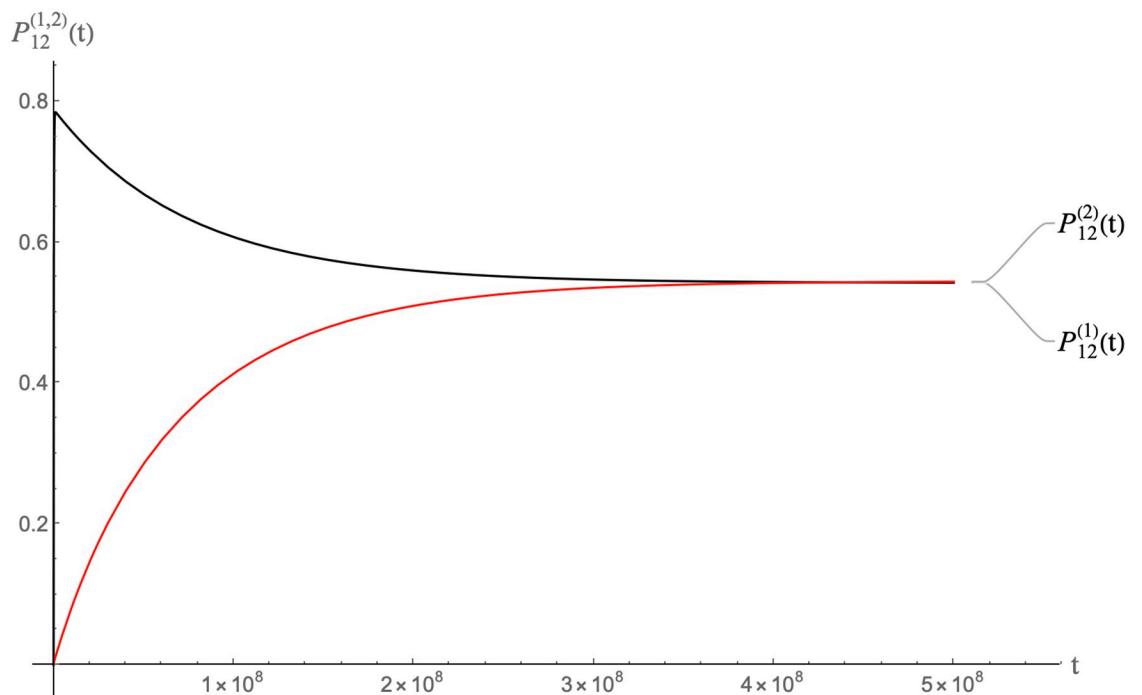


Fig 10. Packaging competition between $MLD = 9, N_p = 8$ spanning trees and $MLD = 19, N_p = 2$ spanning trees with the same parameters as the previous figure on a time scale of 10^8 units.

<https://doi.org/10.1371/journal.pcbi.1009913.g010>

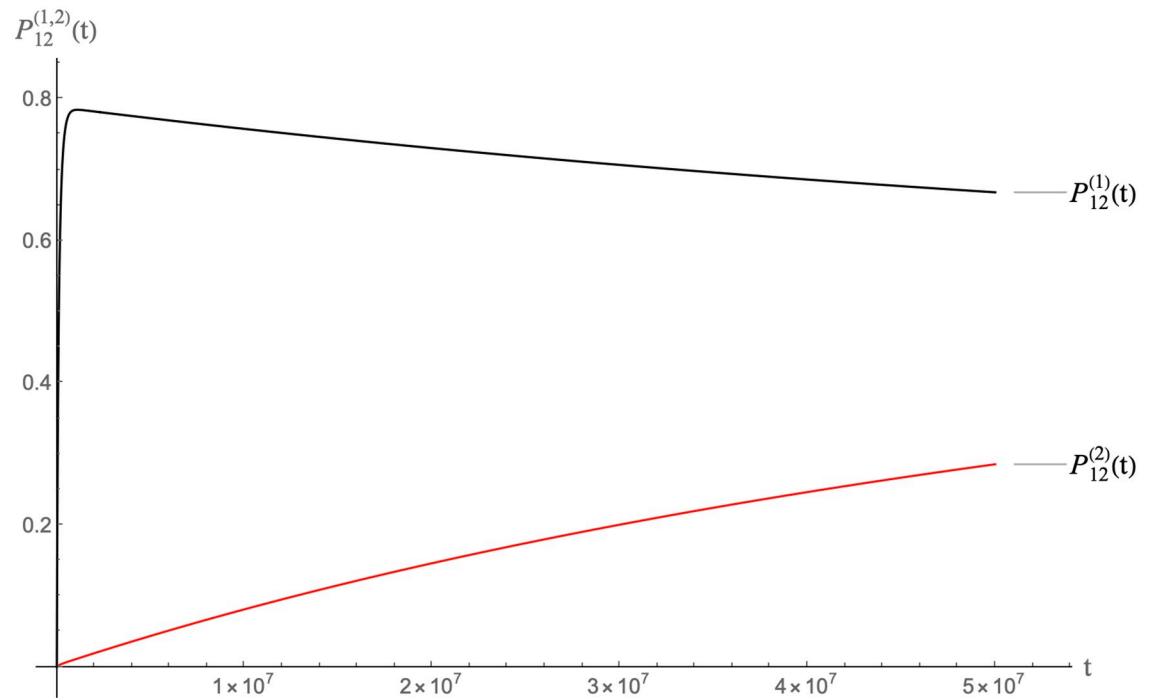


Fig 11. Same as the previous figure but on a time scale of 10^7 units.

<https://doi.org/10.1371/journal.pcbi.1009913.g011>

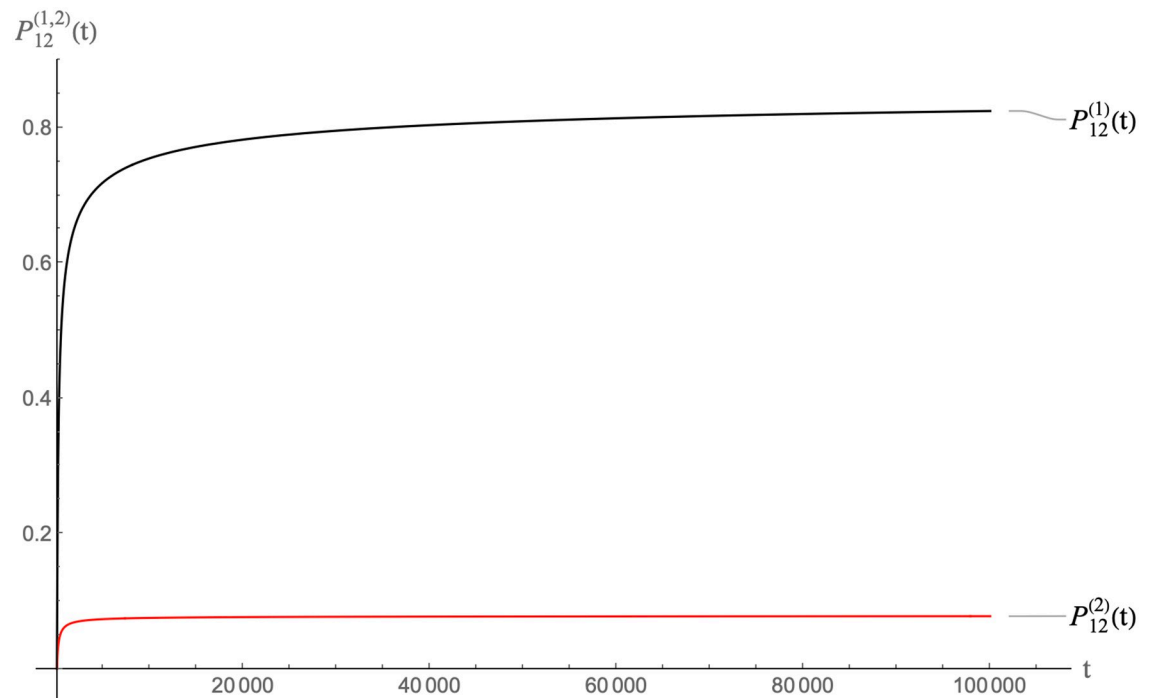


Fig 12. Same as the previous figure but on a time scale of 10^4 units.

<https://doi.org/10.1371/journal.pcbi.1009913.g012>

E_0 ? We will quantify the kinetic selectivity of a packaging competition between class (1) spanning trees that have a small MLD and a large wrapping number and class (2) spanning trees with large MLD and small wrapping number as

$$S(E_0) = \frac{P_{12}^{(1)}(t_{max}) - P_{12}^{(2)}(t_{max})}{P_{12}^{(1)}(t_{max})}. \quad (6)$$

Here, t_{max} is the time at which class (1) spanning trees achieve their maximum packaging yield. If class (2) molecules outcompete class (1) then we set $S(E_0) = 0$. The dependence of the kinetic selectivity on the energy scale on E_0 is shown in Fig 13. The parameter values are the same as those of Fig 10. For E_0 less than about $1.2k_bT$ there is no kinetic selectivity left while $S(E_0)$ approaches one for $E_0 \simeq 4.0$ or larger. The disappearance of selectivity for small E_0 is due to the fact that both the number of assembly pathways and the number of nodes for assembly intermediates of the MLD = 19, $N_p = 2$ trees is two orders of magnitude larger than that of MLD = 9, $N_p = 8$ spanning trees (see Fig 6). This means that for MLD = 19, $N_p = 2$ trees the entropic component of the *free* energy activation barrier causes that barrier to be reduced by a factor of about four to five k_bT . Since the activation energy barrier is about $2E_0$ higher for the MLD = 19, $N_p = 2$ trees, we indeed should expect kinetic selectivity to become negligible for E_0 less than roughly $2k_bT$, in agreement with Fig 13. The importance of entropy for smaller E_0 has another aspect. For E_0 less than about $1.5k_bT$ there is a significant fraction of *incomplete* assemblies since that also increases the entropy of the system. In the language of statistical physics, the kinetic selectivity resembles the order parameter of an order-disorder phase-transition with the energy scale E_0 acting as the inverse of the effective temperature.

One encounters a somewhat similar transition for larger E_0 when one increases the affinity ratio $-\epsilon_1$ between specific pentamer/RNA affinity and pentamer/pentamer affinity. One might expect that that should improve selectivity but, in actuality, for $-\epsilon_1 = -1.1$ a variety of incomplete particles appear packaging MLD = 9, $N_p = 8$ spanning trees. These incomplete particles are in coexistence with fully packaged particles containing MLD = 19, $N_p = 2$ trees. The reason is shown in Fig 14. The minimum energy state at assembly equilibrium of MLD = 9, $N_p = 8$ spanning trees are incomplete particles with ten pentamers while for MLD = 19, $N_p = 2$ trees fully assembled particles are the minimum energy state. For $-\epsilon_1$ smaller than 1.0, the minimum energy state of MLD = 9, $N_p = 8$ spanning trees are fully packaged particles. However metastable states with n less than twelve start to appear roughly above $-\epsilon_1 = 0.6$. Metastable intermediate states are quite familiar from experimental studies of viral assembly [4, 62, 63] and from numerical simulations [13, 36–38]. Known as “kinetic traps”, they retard assembly. In summary, It is clear that increasing $-\epsilon_1$ beyond about 0.5 also does not improve selectivity.

Supersaturation

A different approach to reduce the assembly time scale is to increase the level of supersaturation. Recall in this context that assembly experiments under in-vitro conditions take place at relatively high levels of supersaturation. The supersaturation can be increased by increasing the total pentamer concentration c_0 , which increases the chemical potential by $k_bT \ln c_0$. In Fig 15 the parameters are the same as in Fig 10 except that the pentamer concentration has been increased from $c_0 = 1$ to $c_0 = 4$. The results look encouraging. First, nearly all MLD = 9, $N_p = 8$ spanning trees are packaged. Second, the assembly time scale is reduced to about 10^3 time units, or about one hour. Increasing the supersaturation further reduces assembly time scales. However, the kinetic selectivity *also* has been reduced: the concentration of $N_p = 2$, MLD = 19 assemblies *also* rises much more rapidly than for $c_0 = 1$ under assembly equilibrium

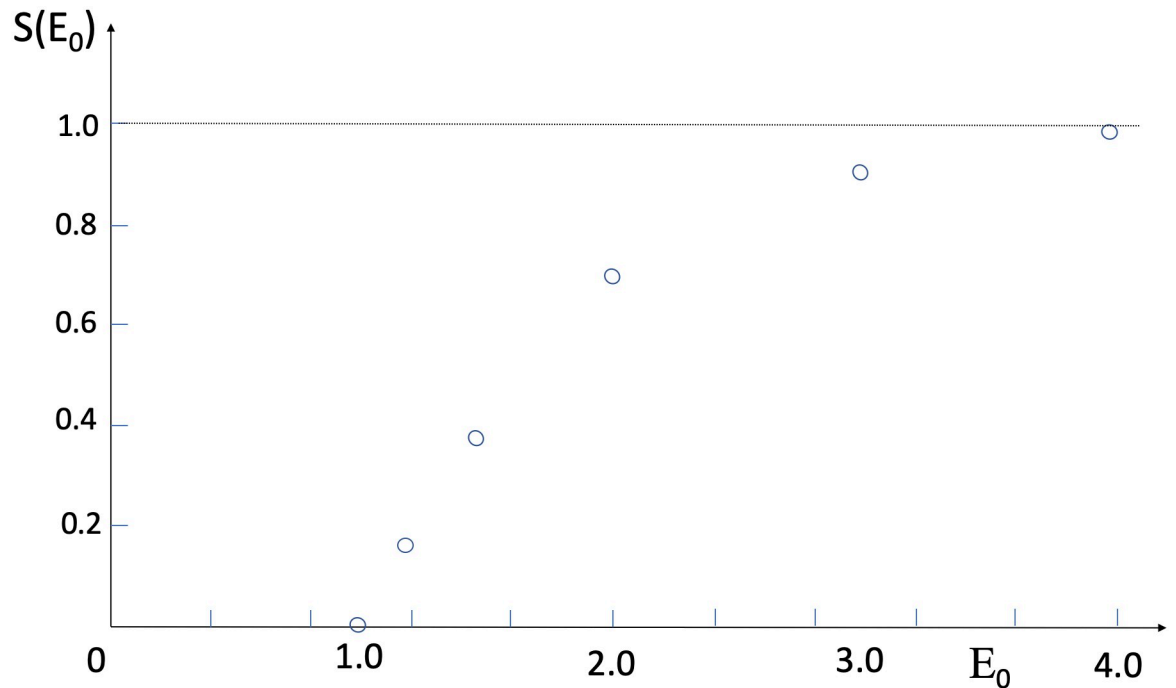


Fig 13. Kinetic selectivity $S(E_0)$ as a function of the energy scale E_0 in units of k_bT for packaging competition between $MLD = 9, N_p = 8$ spanning trees and $MLD = 19, N_p = 2$ spanning trees. The other parameter values are the same as those of Fig 10.

<https://doi.org/10.1371/journal.pcbi.1009913.g013>

conditions. The two classes of molecules have comparable packaging probabilities already after about 10^4 time units.

Is it possible to maintain a high selectivity under conditions of supersaturation that persists over very long time scales? So far we kept the mixing ratio at $D = 0.5$, meaning that the number of pentamers is double of what necessary to package all $MLD = 9, N_p = 8$ and all $N_p = 2, MLD = 19$ spanning trees. What would happen if, under conditions of supersaturation, the

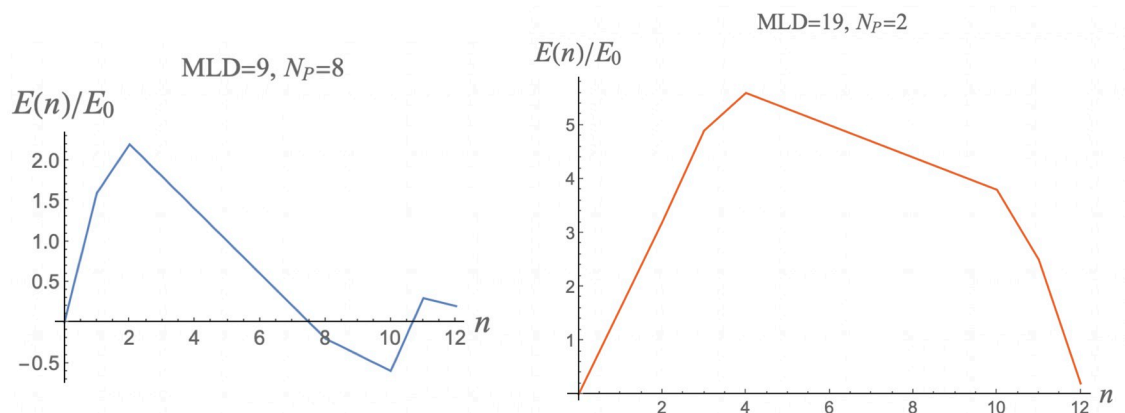


Fig 14. Minimum-energy assembly profiles for $N_p = 8, MLD = 9$ spanning trees (left) and for $N_p = 2, MLD = 19$ spanning trees (right) for $-\epsilon_1 = -1.1$.

<https://doi.org/10.1371/journal.pcbi.1009913.g014>

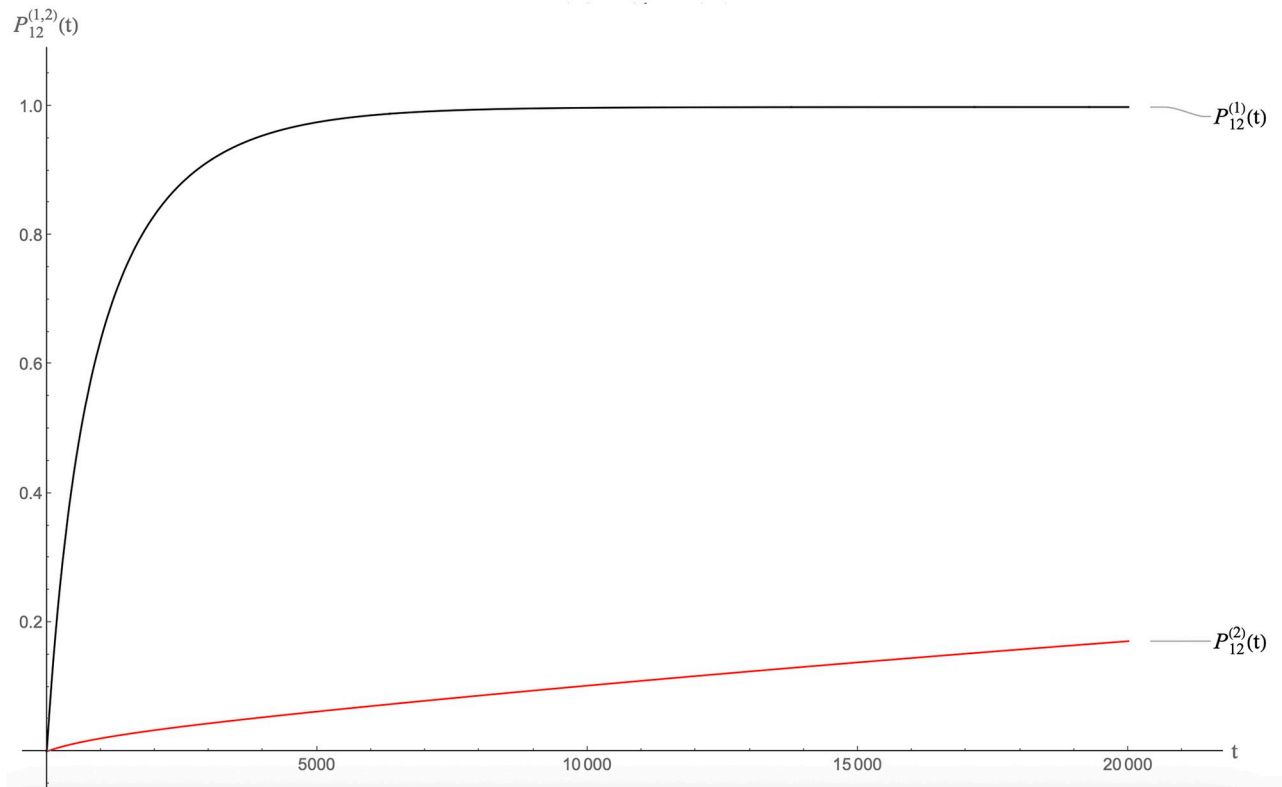


Fig 15. Packaging competition for $c_0 = 4.0$. The other parameters are the same as for Fig 10.

<https://doi.org/10.1371/journal.pcbi.1009913.g015>

mixing ratio would be increased to $D = 2$? In that case, the early assembly of $MLD = 9$, $N_p = 8$ spanning trees might “drain” the solution of pentamers, which would delay the assembly of $N_p = 2$, $MLD = 19$ spanning trees. Fig 16 shows what happens. The kinetic selectivity is about 99 percent and it is maintained over more than 2×10^5 time units! It comes at a price however: the fraction of packaged target molecules is reduced to about 80 percent while the assembly time-scale has mildly increased. In general, by varying the overall pentamer concentration c_0 and the mixing ratio D a wide range of requirements can be satisfied in terms of persistent kinetic selectivity, overall yield, and assembly time-scale.

Parameter values

Are the energy and concentration parameters settings used in this article reasonable for in-vitro or in-vivo viral assembly experiments? The overall energy scale E_0 was defined as the binding affinity between two pentamers that share an edge for the case that the RNA molecule has only generic affinity (i.e. $\epsilon_1 = 0$). The assembly energy per capsid protein of empty capsids has been measured under conditions of thermodynamic equilibrium [64]. Comparing such data to the model with $\epsilon_1 = 0$ gives $E_0 \simeq 4.73$ in units of $k_b T$ [50], close to the value $E_0 = 4$ used in the paper. The other important energy scale is the dimensionless parameter $-\epsilon_1$, which is the ratio between the protein/RNA and the protein-protein affinity. It can be estimated by comparing the capsid protein concentrations at assembly onset for empty capsids and for virions. MS2 capsid proteins aggregate in RNA-free physiological solutions for concentrations above 2.0 mg/ml while in the presence of viral RNA (but not generic RNA) viral particles form

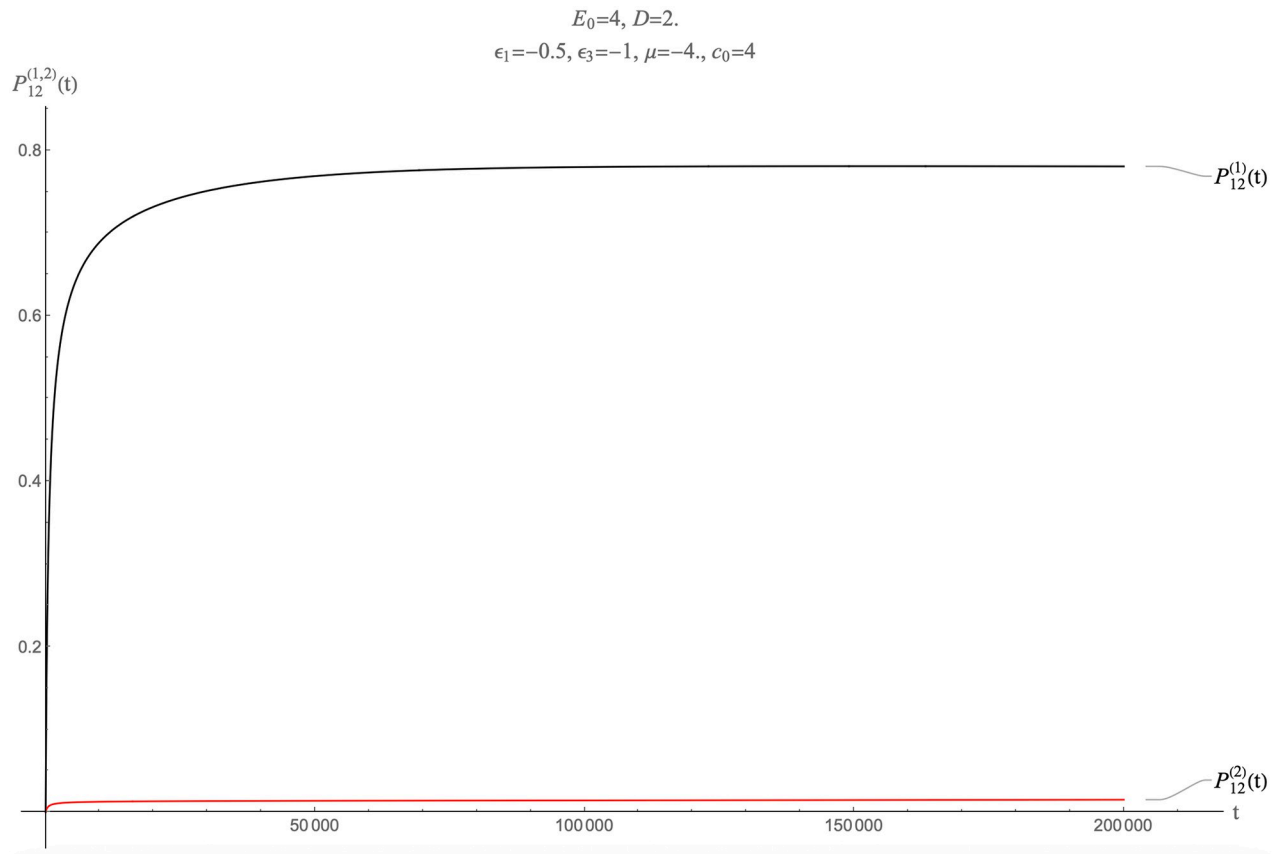


Fig 16. Packaging competition for $c_0 = 4.0$ and $D = 2$. The other parameters are the same as for Fig 10.

<https://doi.org/10.1371/journal.pcbi.1009913.g016>

already at 0.05 mg/ml [65]. The assembly free energy of a viral particle $E_0(-30 + 38\epsilon_1 - 12\mu_0)$ at the reference concentration was set to zero by definition, in which case the empty capsid assembly energy is $-30 - 12\mu_0$ is positive. An increase in the total pentamer concentration by a factor of $2.0/0.05 = 40$ must be able to raise the chemical potential sufficiently so the assembly energy for an empty capsid becomes zero. An increase of the chemical potential equals $\ln 40$ in units of $k_B T$ or about 0.92 in units of $E_0 = 4k_B T$. This condition is satisfied if $-38\epsilon_1 = 12 \times 0.92$ or $-\epsilon_1 \simeq 0.29$. For convenience we used $\epsilon_1 = -0.5$ in the calculations.

Conclusion

In conclusion, we have presented a statistical-mechanics model for the selection of viral ssRNA molecules triggered by packaging signals during the assembly of small RNA viruses. According to the model, there are two important time scales: the characteristic time scale for assembly and the characteristic time scale for thermal equilibration. In the model, RNA selectivity is a non-equilibrium, kinetic effect that disappears when the system approaches thermodynamic equilibrium. Particle *disassembly* under the action of thermal fluctuations is an essential step for thermal equilibration during packaging competition. Kinetic selection “works” as long as the time scale for spontaneous disassembly is prohibitively long compared to the measurement time. Kinetic selection is the result of the dependence of the height of the activation energy barrier on the RNA folding geometry, as encoded by the wrapping number

N_p and the maximum ladder distance MLD. There is an order-disorder type transition as a function of the strength of the affinity between the capsid proteins where fully packaged particles are replaced by a polydisperse solution of incomplete particles. Near the transition, RNA molecules that optimally encode secondary structure and folding geometry—by having minimal MLD and maximal N_p —begin to outcompete other RNA molecules that have a larger number of assembly pathways. A similar order-disorder transition is encountered when the strength of the protein-RNA interaction is increased with respect to that of the protein-protein interactions.

The most striking result is the fact that kinetic selection performs better under increasing levels of supersaturation. There are in fact numerous examples in molecular biology where the fidelity of the read-out of a code is enhanced under non-equilibrium conditions, known as *kinetic proofreading* [66]. DNA replication is an important example [67, 68]. There are similarities between the kinetic selection under supersaturation discussed in this paper and Hopfield kinetic proof reading. Kinetic selection works when the formation of the encoded system (MLD = 9, N_p = 8 spanning tree particles) is quasi-irreversible while attempts of forming particles with molecules that are not encoded (MLD = 19, N_p = 2 spanning tree particles) lead to disassembly. This is the case because of the higher assembly activation energy. An unusual feature is that by tuning the mixing ratio to be close to the stoichiometric ratio for the early assembling encoded particles can “monopolize” the pentamers and thereby greatly retard the formation of improper particles.

An important question about the model concerns the relation between the kinetic selectivity discussed in this paper and selection in terms of the thermodynamic assembly free energy as discussed in the literature cited in the introduction. The relation between the two forms of selection lies in the distinction between the nucleation and elongation stages. An ssRNA molecule may well have packaging signals that produce an unusually low assembly energy barrier causing it to be selected during the nucleation stage. If, however, the molecular weight of the molecule is too high and/or if the solution radius of gyration is too large then the elongation process simply would not be able to complete. It is this elongation part of the assembly that was captured by the earlier studies of the thermodynamic assembly free energy and that was not included in the present study. Following ref. [45], we are assuming here that the selective effects of the packaging signals operate nearly exclusively during the nucleation stage and are weak during the energetically downhill elongation stage which is dominated by the non-specific interactions. The model presented in this paper can be generalized to investigate the competition between selection by nucleation and selection by elongation. The simplest step would be by letting the competing molecules have different reference chemical potentials μ_0 . This allows for the possibility that the two RNA condensates have a different molecular mass and/or a different overall MLDs (recall that the MLD in this paper only refers to the ψ sequence). An additional refinement would be to allow μ_0 to depend on n . As an assembly intermediate grows, an RNA molecule that is only partially condensed will be progressively confined by the developing capsid. This reduces the RNA conformational entropy, which reduces the free energy gain obtained when a pentamer is taken from the solution and added to an incomplete assembly. Finally, the loss of conformational entropy of individual RNA nucleotides as they are getting packaged can be included in the model by associating a fixed amount of entropy with every RNA segment that has not yet associated with a pentamer. This amounts to a renormalization of the ϵ_1 and ϵ_2 parameters.

A second concern with the model is that it is generally assumed that virions are in a state of full thermodynamic equilibrium. If that were truly the case then equilibrium thermodynamics rather than kinetics would *always* be the appropriate description mode. We would argue against the assumption of full thermodynamic equilibrium for virions. Reaching complete

thermodynamic equilibrium in a packaging competition experiment necessarily involves disassembly. The time-dependent assembly yield of early-assembly kinetically encoded molecules needs to decrease in order for thermal equilibrium to be established. Experimentally, the time scale for spontaneous disassembly of virions under the action of thermal fluctuations must be extremely large compared to observation times. We know this because under in-vivo conditions virions typically are released after assembly into environments with few or no capsid proteins. Under those conditions, the chemical potential of the capsid proteins of a virion would be large compared to that of capsid proteins free in solution, which would trigger disassembly. In actuality, spontaneous dissolution of virions in solutions without capsid proteins is not observed under physiological conditions. The time scale for spontaneous disassembly of virions must be large compared to both laboratory time scales and the characteristic time scales of the life cycle of a virus. Maturation processes [63, 69–77] probably play an important role in suppressing spontaneous disassembly process. Maturation stabilization of complete assemblies would further justify a focus on selection during nucleation, as opposed to selection by late-time disassembly/assembly competition necessary for thermalization.

A third concern is that the model presented in this paper is not a realistic representation of any particular virus. It was constructed, for mathematical convenience, by borrowing features of the dodecahedral Zlotnick model for empty capsids, the dodecahedral gRNA spatial distribution of the *Nodaviridae*, and the asymmetric reconstruction of MS2. The wrapping number concept as a geometric characteristic of the geometry of the RNA outer surface really is appropriate only for the special case of *Nodaviridae*. Different geometrical characteristics will have to be developed for other viruses. The MS2 virus is an interesting target since a detailed asymmetric reconstruction of the MS2 virion is available [44]. Since the packaging signals of MS2 gRNA associate directly with capsid proteins—rather than with capsomer edges—the spanning tree would have to have icosahedral rather than dodecahedral symmetry. Moreover, MS2 has (approximate) $T = 3$ capsid symmetry with 180 capsid proteins rather than the $T = 1$ structure with 60 proteins grouped in 12 pentamers that we assumed so that would have to be included as well. We hope to carry out such a study in the future.

A final concern about the model is the study by Tresset and collaborators of the assembly of the CCMV plant virus [78, 79]. They find that the so-called *en-masse* assembly scenario provides a good description [80]. In that scenario, a virion does not assemble on a protein-by-protein basis, as was assumed in the present paper. Instead, assembly starts with the formation of a disordered RNA-protein condensate that shrinks and then transits into an ordered virion. It is tempting to identify CCMV-type assembly (and the *en-masse* assembly scenario) with the entropy-dominated low selectivity assembly scenario that we encountered for lower E_0 and for larger values of the RNA/pentamer affinity. This also will have to be explored in the future.

How could the model (or one of its generalizations) be tested experimentally? It has been shown that large ssRNA molecules in solution with identical primary sequences adopt a range of secondary and tertiary structures [81]. In the presence of a sufficient concentration of positive polyvalent counter ions, large ssRNA molecules have been shown to condense [82]. For a solution of gRNA molecules, that should produce a variety of pre-condensed molecules with roughly the same size but with different surface structures. Asymmetric reconstruction of the packaged particles could then reveal which RNA structures were selected for.

Supporting information

S1 Text. Smoluchowski theory of bimolecular reactions.
(PDF)

Acknowledgments

We would like to thank Alexander Grosberg for suggesting the spanning tree, Ioulia Rouzina for suggesting selective nucleation as a central mechanism of viral assembly, Chuck Knobler and Reidun Twarock for reading a draft of the manuscript and Ioulia Rouzina, Orlando Guzman, Chuck Knobler, William Gelbart, Chen Lin, Zach Gvildys and William Vong for helpful discussions.

Author Contributions

Conceptualization: Inbal Mizrahi, Robijn Bruinsma, Joseph Rudnick.

Formal analysis: Robijn Bruinsma, Joseph Rudnick.

Investigation: Inbal Mizrahi, Joseph Rudnick.

Methodology: Joseph Rudnick.

Resources: Robijn Bruinsma.

Supervision: Robijn Bruinsma, Joseph Rudnick.

Validation: Joseph Rudnick.

Visualization: Inbal Mizrahi.

Writing – original draft: Robijn Bruinsma, Joseph Rudnick.

Writing – review & editing: Inbal Mizrahi, Robijn Bruinsma, Joseph Rudnick.

References

1. Dimmock NJ, Easton AJ, Leppard KN. Introduction to modern virology. Malden, MA: Blackwell Publishing; 2001.
2. Frensing T, Kupke SY, Bachmann M, Fritzsche S, Gallo-Ramirez LE, Reichl U. Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in MDCK cells. *Applied microbiology and biotechnology*. 2016; 100(16):7181–7192. <https://doi.org/10.1007/s00253-016-7542-4> PMID: 27129532
3. Frolova E, Frolov I, Schlesinger S. Packaging signals in alphaviruses. *J Virol*. 1997; 71(1):248–258. <https://doi.org/10.1128/jvi.71.1.248-258.1997> PMID: 8985344
4. Basnak G, Morton VL, Rolfsson O, Stonehouse NJ, Ashcroft AE, Stockley PG. Viral Genomic Single-Stranded RNA Directs the Pathway Toward a T = 3 Capsid. *J Mol Biol*. 2010; 395(5):924–936. <https://doi.org/10.1016/j.jmb.2009.11.018> PMID: 19913556
5. Bunka DHJ, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, et al. Degenerate RNA Packaging Signals in the Genome of Satellite Tobacco Necrosis Virus: Implications for the Assembly of a T = 1 Capsid. *J Mol Biol*. 2011; 413(1):51–65. <https://doi.org/10.1016/j.jmb.2011.07.063> PMID: 21839093
6. Stockley PG, Twarock R, Bakker SE, Barker AM, Borodavka A, Dykeman E, et al. Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J Biol Phys*. 2013; 39(2):277–287. <https://doi.org/10.1007/s10867-013-9313-0> PMID: 23704797
7. Dykeman EC, Stockley PG, Twarock R. Building a Viral Capsid in the Presence of Genomic RNA. *Phys Rev E*. 2013; 87(2):022717. <https://doi.org/10.1103/PhysRevE.87.022717> PMID: 23496558
8. Dykeman EC, Stockley PG, 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):3235–49. <https://doi.org/10.1016/j.jmb.2013.06.005> PMID: 23763992
9. Patel N, Dykeman EC, Coutts RHA, Lomonosoff GP, Rowlands DJ, Phillips SEV, et al. Revealing the density of encoded functions in a viral RNA. *Proc Natl Acad Sci USA*. 2015; <https://doi.org/10.1073/pnas.1420812112> PMID: 25646435
10. van der Schoot P, Bruinsma R. Electrostatics and the assembly of an RNA virus. *Phys Rev E*. 2005; 71(6):061928. <https://doi.org/10.1103/PhysRevE.71.061928> PMID: 16089786
11. Forrey C, Muthukumar M. Electrostatics of capsid-induced viral RNA organization. *J Chem Phys*. 2009; 131(10). <https://doi.org/10.1063/1.3216550>

12. Garmann RF, Comas-Garcia M, Koay MST, Cornelissen JJLM, Knobler CM, Gelbart WM. The Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus. *J Virol*. 2014; <https://doi.org/10.1128/JVI.01044-14> PMID: 24965458
13. Perlmutter JD, Hagan MF. The Role of Packaging Sites in Efficient and Specific Virus Assembly. *J Mol Biol*. 2015; <https://doi.org/10.1016/j.jmb.2015.05.008> PMID: 25986309
14. Prevelige PE, Thomas D, King J. Nucleation and Growth Phases in the Polymerization of Coat and Scaffolding Subunits into Icosahedral Procapsid Shells. *Biophys J*. 1993; 64(3):824–835. [https://doi.org/10.1016/S0006-3495\(93\)81443-7](https://doi.org/10.1016/S0006-3495(93)81443-7) PMID: 8471727
15. Casini GL, Graham D, Heine D, Garcea RL, Wu DT. In Vitro Papillomavirus Capsid Assembly Analyzed by Light Scattering. *Virology*. 2004; 325(2):320–327. <https://doi.org/10.1016/j.virol.2004.04.034> PMID: 15246271
16. Medrano M, Fuertes MÁ, Valbuena A, Carrillo PJ, Rodríguez-Huete A, Mateu MG. Imaging and quantitation of a succession of transient intermediates reveal the reversible self-assembly pathway of a simple icosahedral virus capsid. *Journal of the American Chemical Society*. 2016; 138(47):15385–15396. <https://doi.org/10.1021/jacs.6b07663> PMID: 27933931
17. Zandi R, van der Schoot P, Reguera D, Kegel W, Reiss H. Classical Nucleation Theory of Virus Capsids. *Biophys J*. 2006; 90(6):1939–1948. <https://doi.org/10.1529/biophysj.105.072975> PMID: 16387781
18. Bruinsma RF, Wuite GJ, Roos WH. Physics of viral dynamics. *Nature Reviews Physics*. 2021; p. 1–16. <https://doi.org/10.1038/s42254-020-00267-1> PMID: 33728406
19. Zhang DQ, Konecny R, Baker NA, McCammon JA. Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers*. 2004; 75(4):325–337. <https://doi.org/10.1002/bip.20120> PMID: 15386271
20. Kegel WK, van der Schoot P. Physical regulation of the self-assembly of tobacco mosaic virus coat protein. *Biophys J*. 2006; 91(4):1501–1512. <https://doi.org/10.1529/biophysj.105.072603> PMID: 16731551
21. Belyi VA, Muthukumar M. Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci U S A*. 2006; 103(46):17174–17178. <https://doi.org/10.1073/pnas.0608311103> PMID: 17090672
22. Hu T, Zhang R, Shklovskii BI. Electrostatic theory of viral self-assembly. *Physica A*. 2008; 387(12):3059–3064. <https://doi.org/10.1016/j.physa.2008.01.010>
23. Devkota B, Petrov AS, Lemieux S, Boz MB, Tang L, Schneemann A, et al. Structural and Electrostatic Characterization of Pariacoto Virus: Implications for Viral Assembly. *Biopolymers*. 2009; 91(7):530–538. <https://doi.org/10.1002/bip.21168> PMID: 19226622
24. Hagan MF. A Theory for Viral Capsid Assembly around Electrostatic Cores. *J Chem Phys*. 2009; 130:114902. <https://doi.org/10.1063/1.3086041> PMID: 19317561
25. Jiang T, Wang ZG, Wu JZ. Electrostatic Regulation of Genome Packaging in Human Hepatitis B Virus. *Biophys J*. 2009; 96(8):3065–3073. <https://doi.org/10.1016/j.bpj.2009.01.009> PMID: 19383452
26. Siber A, Zandi R, Podgornik R. Thermodynamics of nanospheres encapsulated in virus capsids. *Phys Rev E*. 2010; 81(5):051919. <https://doi.org/10.1103/PhysRevE.81.051919> PMID: 20866273
27. Ting CL, Wu J, Wang ZG. Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proc Natl Acad Sci U S A*. 2011; 108(41):16986–16991. <https://doi.org/10.1073/pnas.1109307108> PMID: 21969546
28. Ni P, Wang Z, Ma X, Das NC, Sokol P, Chiu W, et al. An Examination of the Electrostatic Interactions between the N-Terminal Tail of the Coat Protein and RNA in Brome Mosaic Virus. *J Mol Biol*. 2012; 419:284–300. <https://doi.org/10.1016/j.jmb.2012.03.023> PMID: 22472420
29. Siber A, Bozic AL, Podgornik R. Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Phys Chem Chem Phys*. 2012; 14(11):3746–3765. <https://doi.org/10.1039/C1CP22756D> PMID: 22143065
30. Ford RJ, Barker AM, Bakker SE, Coutts RH, Ranson NA, Phillips SE, et al. Sequence-specific, RNA–protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J Mol Biol*. 2013; 425(6):1050–1064. <https://doi.org/10.1016/j.jmb.2013.01.004> PMID: 23318955
31. Zhang R, Linse P. Icosahedral capsid formation by capsomers and short polyions. *J Chem Phys*. 2013; 138(15). <https://doi.org/10.1063/1.4799243> PMID: 23614442
32. Erdemci-Tandogan G, Wagner J, van der Schoot P, Podgornik R, Zandi R. RNA topology remodels electrostatic stabilization of viruses. *Phys Rev E*. 2014; 89:032707. <https://doi.org/10.1103/PhysRevE.89.032707> PMID: 24730874
33. Garmann RF, Comas-Garcia M, Koay MST, Cornelissen J, Knobler CM, Gelbart WM. Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus. *J Virol*. 2014; 88(18):10472–10479. <https://doi.org/10.1128/JVI.01044-14> PMID: 24965458

34. Kim J, Wu JZ. A Thermodynamic Model for Genome Packaging in Hepatitis B Virus. *Biophys J*. 2015; 109(8):1689–1697. <https://doi.org/10.1016/j.bpj.2015.08.021> PMID: 26488660
35. Bond K, Tsvetkova IB, Wang JCY, Jarrold MF, Dragnea B. Virus Assembly Pathways: Straying Away but Not Too Far. *Small*. 2020; 16(51):2004475. <https://doi.org/10.1002/smll.202004475>
36. Johnston IG, Louis AA, Doye JPK. Modelling the self-assembly of virus capsids. *J Phys: Condens Matter*. 2010; 22(10):104101. PMID: 21389435
37. Hagan MF, Elrad OM, Jack RL. Mechanisms of Kinetic Trapping in Self-Assembly and Phase Transformation. *J Chem Phys*. 2011; 135:104115. <https://doi.org/10.1063/1.3635775> PMID: 21932884
38. Baschek JE, Klein HCR, Schwarz US. Stochastic dynamics of virus capsid formation: direct versus hierarchical self-assembly. *Bmc Biophysics*. 2012; 5. <https://doi.org/10.1186/2046-1682-5-22> PMID: 23244740
39. Roos W, Bruinsma R, Wuite G. Physical virology. *Nature physics*. 2010; 6(10):733–743. <https://doi.org/10.1038/nphys1797>
40. Zandi R, Dragnea B, Travasset A, Podgornik R. On virus growth and form. *Physics Reports*. 2020; 847:1–102. <https://doi.org/10.1016/j.physrep.2019.12.005>
41. Klug A. The Tobacco Mosaic Virus Particle: Structure and Assembly. *Philos Trans R Soc Lond B Biol Sci*. 1999; 354(1383):531–535. <https://doi.org/10.1098/rstb.1999.0404> PMID: 10212932
42. Comas-Garcia M, Datta SA, Baker L, Varma R, Gudla PR, Rein A. Dissection of specific binding of HIV-1 Gag to the packaging signal in viral RNA. *Elife*. 2017; 6. <https://doi.org/10.7554/eLife.27055> PMID: 28726630
43. Jouvenet N, Simon SM, Bieniasz PD. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proceedings of the National Academy of Sciences*. 2009; 106(45):19114–19119. <https://doi.org/10.1073/pnas.0907364106> PMID: 19861549
44. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, et al. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nature communications*. 2016; 7(1):1–6. <https://doi.org/10.1038/ncomms12524> PMID: 27561669
45. Dykeman EC, Grayson NE, Toropova K, Ranson NA, Stockley PG, Twarock R. Simple Rules for Efficient Assembly Predict the Layout of a Packaged Viral RNA. *J Mol Biol*. 2011; 408(3):399–407. <https://doi.org/10.1016/j.jmb.2011.02.039> PMID: 21354423
46. Beren C, Cui Y, Chakravarty A, Yang X, Rao A, Knobler CM, et al. Genome organization and interaction with capsid protein in a multipartite RNA virus. *Proceedings of the National Academy of Sciences*. 2020; 117(20):10673–10680. <https://doi.org/10.1073/pnas.1915078117>
47. Cadena-Nava RD, Hu Y, Garmann RF, Ng B, Zelikin AN, Knobler CM, et al. Exploiting fluorescent polymers to probe the self-assembly of virus-like particles. *J Phys Chem B*. 2011; 115(10):2386–91. <https://doi.org/10.1021/jp1094118> PMID: 21338131
48. Cadena-Nava RD, Comas-Garcia M, Garmann RF, Rao A, Knobler CM, Gelbart WM. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *Journal of virology*. 2012; 86(6):3318–3326. <https://doi.org/10.1128/JVI.06566-11> PMID: 22205731
49. Tang L, Johnson KN, Ball LA, Lin T, Yeager M, Johnson JE. The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nature Structural & Molecular Biology*. 2001; 8(1):77–83. <https://doi.org/10.1038/83089> PMID: 11135676
50. Zlotnick A. To Build a Virus Capsid—an Equilibrium-Model of the Self-Assembly of Polyhedral Protein Complexes. *J Mol Biol*. 1994; 241(1):59–67. <https://doi.org/10.1006/jmbi.1994.1473> PMID: 8051707
51. Endres D, Zlotnick A. Model-Based Analysis of Assembly Kinetics for Virus Capsids or Other Spherical Polymers. *Biophys J*. 2002; 83(2):1217–1230. [https://doi.org/10.1016/S0006-3495\(02\)75245-4](https://doi.org/10.1016/S0006-3495(02)75245-4) PMID: 12124301
52. Zlotnick A. Distinguishing Reversible from Irreversible Virus Capsid Assembly. *J Mol Biol*. 2007; 366(1):14–18. <https://doi.org/10.1016/j.jmb.2006.11.034> PMID: 17157314
53. Morozov AY, Bruinsma RF, Rudnick J. Assembly of viruses and the pseudo-law of mass action. *J Chem Phys*. 2009; 131(15):155101. <https://doi.org/10.1063/1.3212694> PMID: 20568884
54. Graham RL, Hell P. On the history of the minimum spanning tree problem. *Annals of the History of Computing*. 1985; 7(1):43–57. <https://doi.org/10.1109/MAHC.1985.10011>
55. Rudnick J, Bruinsma R. Icosahedral packing of RNA viral genomes. *Phys Rev Lett*. 2005; 94(3):038101. <https://doi.org/10.1103/PhysRevLett.94.038101> PMID: 15698326
56. Twarock R, Leonov G, Stockley PG. Hamiltonian path analysis of viral genomes. *Nature communications*. 2018; 9(1):1–3. <https://doi.org/10.1038/s41467-018-03713-y> PMID: 29789600

57. Yoffe AM, Prinsen P, Gopal A, Knobler CM, Gelbart WM, Ben-Shaul A. Predicting the sizes of large RNA molecules. *Proceedings of the National Academy of Sciences*. 2008; 105(42):16153–16158. <https://doi.org/10.1073/pnas.0808089105> PMID: 18845685
58. Fang LT, Gelbart WM, Ben-Shaul A. The size of RNA as an ideal branched polymer. *The Journal of Chemical Physics*. 2011; 135(15):10B616. <https://doi.org/10.1063/1.3652763> PMID: 22029339
59. Bollobás B. *Modern graph theory*. vol. 184. Springer Science & Business Media; 2013.
60. Van Kampen NG. *Stochastic processes in physics and chemistry*. vol. 1. Elsevier; 1992.
61. Schulten K. *Non-equilibrium Statistical Mechanics*. UIUC; 1999.
62. Parent KN, Zlotnick A, Teschke CM. Quantitative analysis of multi-component spherical virus assembly: Scaffolding protein contributes to the global stability of phage P22 procapsids. *J Mol Biol*. 2006; 359(4):1097–1106. <https://doi.org/10.1016/j.jmb.2006.03.068> PMID: 16697406
63. Tuma R, Tsuruta H, French KH, Prevelige PE. Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid. *J Mol Biol*. 2008; 381(5):1395–1406. <https://doi.org/10.1016/j.jmb.2008.06.020> PMID: 18582476
64. Ceres P, Zlotnick A. Weak Protein-Protein Interactions Are Sufficient to Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry*. 2002; 41(39):11525–11531. <https://doi.org/10.1021/bi0261645> PMID: 12269796
65. Matthews KS, Cole RD. Shell formation by capsid protein of f2 bacteriophage. *Journal of molecular biology*. 1972; 65(1):1–15. [https://doi.org/10.1016/0022-2836\(72\)90487-1](https://doi.org/10.1016/0022-2836(72)90487-1) PMID: 4553257
66. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*. 1974; 71(10):4135–4139. <https://doi.org/10.1073/pnas.71.10.4135> PMID: 4530290
67. Echols H, Goodman MF. Fidelity mechanisms in DNA replication. *Annual review of biochemistry*. 1991; 60(1):477–511. <https://doi.org/10.1146/annurev.bi.60.070191.002401> PMID: 1883202
68. Tippin B, Pham P, Goodman MF. Error-prone replication for better or worse. *Trends in microbiology*. 2004; 12(6):288–295. <https://doi.org/10.1016/j.tim.2004.04.004> PMID: 15165607
69. Garoff H, Hewson R, Opstelten DJE. Virus maturation by budding. *Microbiology and Molecular Biology Reviews*. 1998; 62(4):1171–+. <https://doi.org/10.1128/MMBR.62.4.1171-1190.1998> PMID: 9841669
70. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*. 2000; 289(5487):2129–2133. <https://doi.org/10.1126/science.289.5487.2129> PMID: 11000116
71. Guerin T, Bruinsma R. Theory of conformational transitions of viral shells. *Phys Rev E*. 2007; 76:061911–061911. <https://doi.org/10.1103/PhysRevE.76.061911> PMID: 18233873
72. Parent KN, Suhanovsky MM, Teschke CM. Polyhead formation in phage P22 pinpoints a region in coat protein required for conformational switching. *Mol Microbiol*. 2007; 65(5):1300–1310. <https://doi.org/10.1111/j.1365-2958.2007.05868.x> PMID: 17680786
73. Roos WH, Gertsman I, May ER, Brooks CL, Johnson JE, Wuite GJL. Mechanics of bacteriophage maturation. *Proc Natl Acad Sci U S A*. 2012; 109(7):2342–2347. <https://doi.org/10.1073/pnas.1109590109> PMID: 22308333
74. May ER, Feng J, Brooks CL. Exploring the Symmetry and Mechanism of Virus Capsid Maturation Via an Ensemble of Pathways. *Biophys J*. 2012; 102(3):606–612. <https://doi.org/10.1016/j.bpj.2011.12.016> PMID: 22325284
75. Veesler D, Johnson JE. Virus Maturation. *Annual Review of Biophysics*. 2012; 41(1):473–496. <https://doi.org/10.1146/annurev-biophys-042910-155407> PMID: 22404678
76. Domitrovic T, Movahed N, Bothner B, Matsui T, Wang Q, Doerschuk PC, et al. Virus Assembly and Maturation: Auto-Regulation through Allosteric Molecular Switches. *J Mol Biol*. 2013; 425(9):1488–1496. <https://doi.org/10.1016/j.jmb.2013.02.021> PMID: 23485419
77. Mateu MG. Assembly, stability and dynamics of virus capsids. *Arch Biochem Biophys*. 2013; 531(1–2):65–79. <https://doi.org/10.1016/j.abb.2012.10.015> PMID: 23142681
78. Panahandeh S, Li S, Marichal L, Leite Rubim R, Tresset G, Zandi R. How a virus circumvents energy barriers to form symmetric shells. *ACS nano*. 2020; 14(3):3170–3180. <https://doi.org/10.1021/acsnano.9b08354> PMID: 32115940
79. Chevreuil M, Law-Hine D, Chen J, Bressanelli S, Combet S, Constantin D, et al. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nature communications*. 2018; 9(1):1–9. <https://doi.org/10.1038/s41467-018-05426-8> PMID: 30082710
80. Perlmutter JD, Hagan MF. Mechanisms of Virus Assembly. *Annu Rev Phys Chem*. 2015; 66(1):217–239. <https://doi.org/10.1146/annurev-physchem-040214-121637> PMID: 25532951

81. Garmann RF, Gopal A, Athavale SS, Knobler CM, Gelbart WM, Harvey SC. Visualizing the Global Secondary Structure of a Viral RNA Genome with Cryo-Electron Microscopy. *RNA*. 2015; 21(5):877–886. <https://doi.org/10.1261/rna.047506.114> PMID: 25752599
82. Kapuscinski J, Darzynkiewicz Z. Condensation of nucleic acids by intercalating aromatic cations. *Proceedings of the National Academy of Sciences*. 1984; 81(23):7368–7372. <https://doi.org/10.1073/pnas.81.23.7368> PMID: 6209715