Building a viral capsid in the presence of genomic RNA

Eric C. Dykeman,[1,2] Peter G. Stockley,[3] and Reidun Twarock[1,2]
[1]*Department of Biology, York Centre for Complex Systems Analysis, University of York, York, YO10 5DD United Kingdom*
[2]*Department of Mathematics, York Centre for Complex Systems Analysis, University of York, York, YO10 5DD United Kingdom*
[3]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT United Kingdom*

(Received 19 September 2012; revised manuscript received 29 November 2012; published 25 February 2013)

Virus capsid assembly has traditionally been considered as a process that can be described primarily via self-assembly of the capsid proteins, neglecting interactions with other viral or cellular components. Our recent work on several ssRNA viruses, a major class of viral pathogens containing important human, animal, and plant viruses, has shown that this protein-centric view is too simplistic. Capsid assembly for these viruses relies strongly on a number of cooperative roles played by the genomic RNA. This realization requires a new theoretical framework for the modeling and prediction of the assembly behavior of these viruses. In a seminal paper Zlotnick [J. Mol. Biol. **241**, 59 (1994)] laid the foundations for the modeling of capsid assembly as a protein-only self-assembly process, illustrating his approach using the example of a dodecahedral study system. We describe here a generalized framework for modeling assembly that incorporates the regulatory functions provided by cognate protein–nucleic-acid interactions between capsid proteins and segments of the genomic RNA, called packaging signals, into the model. Using the same dodecahedron system we demonstrate, using a Gillespie-type algorithm to deal with the enhanced complexity of the problem instead of a master equation approach, that assembly kinetics and yield strongly depend on the distribution and nature of the packaging signals, highlighting the importance of the crucial roles of the RNA in this process.

DOI: 10.1103/PhysRevE.87.022717 PACS number(s): 87.16.dr, 87.15.A−, 87.10.Mn

I. INTRODUCTION

The self-assembly of viral capsids from their constituent capsid proteins is one of the prime examples of molecular self-assembly. Traditionally these have been considered as processes determined entirely by the assembly of the capsid proteins, i.e., by neglecting effects due to the genomic RNA or other viral components such as scaffolding proteins [1]. There are two basic mechanisms of capsid assembly used by viruses. In the first, a procapsid lacking nucleic acid is constructed and subsequent work must be done to package the genome into this structure. The second involves spontaneous coassembly of coat proteins and nucleic acid to form the virion. Zlotnick's work was one of the first to introduce a theoretical framework for this [2]. In this approach, the gradual buildup of a viral capsid from its protein building blocks is modeled via a set of kinetic equations that specify the individual reactions occurring between capsid proteins during the process. By solving the corresponding set of differential equations for the concentrations of the assembly intermediates, their concentrations are predicted and thermodynamic as well as kinetic descriptions of this process have been established [2–4].

Recently, we have shown that in several ssRNA viruses that use the capsid coassembly mechanism, a number of contacts between viral coat proteins and sequences within their cognate genomic RNAs, called packaging signals (PSs), play essential roles in ensuring efficient capsid assembly [5–9]. Examples include the bacteriophage MS2 [5,6,10–16] and the plant satellite virus, Satellite tobacco necrosis virus (STNV) [8,17,18]. While assembly in the absence of these contacts is possible *in vitro*, it is mostly very slow and inefficient. Such a process would be nonviable *in vivo* because for most cells and hosts, partially assembled capsids would trigger antiviral defense mechanisms such as an immune response or RNA silencing. We therefore revisit the assembly modeling problem here by discussing ways of incorporating the roles of the genomic RNA.

Although the roles of the strongest packaging signals, e.g., in initiating the assembly process, have long been recognized in the experimental literature [5,12,13,19], the existence and roles of the weaker ones have traditionally been overlooked. This is due to the fact that their structures are defined in terms of general sequence or structure motifs, rather than by repetition of identical contiguous sequence segments, making simple bioinformatics analyses that attempt to locate them in ssRNA genomes unsuccessful [20]. Using a combination of RNA systematic evolution of ligands by exponential enrichment (SELEX), a technique that identifies sequences with affinities for a target protein such as a viral coat protein, structural and biochemical information on RNA recognition, and a new bioinformatics approach, we were recently able to predict the locations of the potential lower-affinity PSs in the genomes of STNV [8] and bacteriophage MS2 [9]. In the latter case this even included predicting their locations in the tertiary structure of the packaged genome. In addition to these theoretical and experimental data, recent assembly experiments have demonstrated the vital roles played by multiple lower-affinity PSs in the cognate genomes of ssRNA viruses in ensuring faithful capsid assembly. These experiments revealed a two-stage assembly process, where in the first stage multiple cognate RNA–coat protein (CP) interactions cause a collapse of the hydrodynamic radius of the RNA so that it will fit within the capsid. This effect is specific to cognate viral RNAs. Noncognate RNAs can be packaged into virus-like particles, but much less efficiently and with lower fidelity than the packaging of the cognate genome [15].

The predictions of lower-affinity PSs in the genomes of STNV and MS2, along with this recent experimental evidence, strongly imply that current models of packaging mechanisms

©2013 American Physical Society

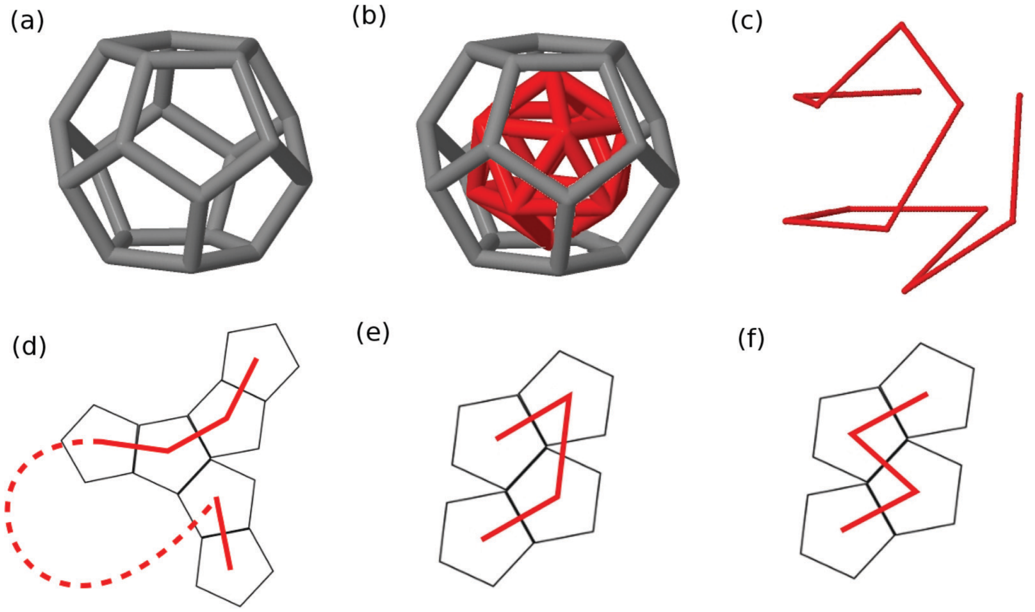

FIG. 1. (Color online) Geometry of the dodecahedron study system. (a) Dodecahedron as a coarse-grained model for a viral capsid formed from 12 pentamers. (b) The RNA-protein contacts at the centers of the pentagonal faces inscribe an icosahedron, here shown in red. (c) Example of a Hamiltonian path on the icosahedron. (d) Example of a protein intermediate that cannot occur during coassembly with RNA. Dashed lines illustrate how this intermediate would violate the local-rule-assembly model by jumping to another part of the shell instead of moving to one of the four neighboring pentamer sites that are vacant. (e) and (f) Two assembly intermediates with identical protein configurations but different RNA organizations. These are considered as different intermediates in the case of assembly in the presence of RNA.

for ssRNA viruses in which the RNA genome is considered as a homogeneous polyelectrolyte [21–25] are not sufficient to account for these observations. Indeed, the experimental data call for a paradigm shift in the modeling of the ssRNA virus assembly, requiring a new theoretical framework that supersedes simple electrostatic models, includes information on the PSs and their varying affinities to coat protein, and is able to reproduce the differences in assembly efficiency observed for cognate and noncognate RNAs. Packaging signals can in principle be incorporated into kinetic models of capsid assembly by introducing additional reactions between RNA and capsid protein. However, as we discuss in the following section, this increases the complexity of the model dramatically, so that techniques reliant on the solution of differential equations, as in Zlotnick's approach [2,4], are no longer viable. We therefore introduce an algorithm that incorporates RNA–coat protein binding events into the assembly model in a way that avoids this complexity problem, akin to complexity reduction in approaches taken in the protein-only cases considered by Schwartz and co-workers [26,27]. This makes our approach scalable to much larger viral systems. Our results demonstrate important insights into the roles of PSs in ssRNA virus assembly. Based on the dodecahedron model system, they illustrate that yield and speed of assembly vary significantly both with the location in the RNA of the strongest packaging signal that initiates assembly and with the overall locations of the other packaging signals in the RNA and their affinity for coat protein. We hence provide a theoretical explanation for the vital roles of lower-affinity PSs in ensuring yield and speed of capsid assembly.

II. MODEL SYSTEM

As a model system we use the dodecahedron, which can be considered as a coarse-grained representation of a virus assembling from 12 clusters of five proteins (pentamers), with RNA binding sites at the centers of the pentamers [see Fig. 1(a)]. Examples of such capsids are found in the Picornaviridae, e.g., Equine rhinitis A virus [28], and a number of small plant viruses of the Comovirus family, e.g., Cowpea mosaic virus [29]. Zlotnick considered assembly of a dodecahedral capsid via sequential attachment of individual pentamers from both an experimental [1] and a theoretical point of view [2]. He described this process via an assembly graph representing all possible pathways to the fully formed capsid that can occur via attachment of a single pentamer at a time. If constraints from interactions with genomic RNA are neglected, pentamers are free to attach at any unoccupied interface on the growing capsid. However, if contacts with genomic RNA must be made, attachment of a capsid protein to the growing protein shell will be constrained by the secondary and tertiary structures of the RNA. As discussed in previous work [30], it is unlikely that, as the genomic RNA and protein shell coassemble, the RNA would suddenly move a large distance from one area of the growing shell to another. In this context, our model assumes that RNA–coat-protein coassembly is governed by local assembly rules, a simpler view that is more consistent with experimental observations such as the inner and outer shells of RNA observed in cryo–electron-microscopy (cryo-EM) reconstructions of bacteriophage MS2 [31] and many other ssRNA viruses. In the local-rule-assembly model [30], CP can only be added next to the two CPs that are bound to the ends of the RNA fragment in complex with the partially formed shell.

For the dodecahedron model system, this implies that addition of incoming CP is only possible at specific sites that are consistent with the RNA forming contacts at all vertices of an inscribed icosahedron [see Fig. 1(b)], i.e., those forming Hamiltonian paths on this icosahedron such as the one shown in Fig. 1(c). As a result, some of the assembly intermediates in the RNA-free case are no longer accessible in the context of a local-rule-coassembly process. An example of an intermediate that is no longer accessible, because there is no connected path between the centers of all its pentagonal faces, is shown in Fig. 1(d). In this illustration the RNA adds the first four sequential coat proteins following a local-rule-assembly model. It then moves to another part of the growing shell, as shown by the dashed line, forming a disconnected path in violation of the model. In contrast, a given protein configuration can occur with different RNA layouts [see, e.g., Figs. 1(e) and 1(f)]. Since these need to be treated as different assembly intermediates, this leads to an increased ensemble of assembly intermediates. Therefore, the complexity of the problem increases when RNA interactions are taken into account. As a result, modeling assembly via kinetic equations and numerically solving the differential equations for the concentrations of the assembly intermediates is no longer practicable here. We therefore introduce an algorithm, based on the Gillespie algorithm [32] and a variation used by Schwartz and co-workers [26,27], to examine an assembly of empty capsids, that is capable of handling the additional complexity that arises from considering coat protein interactions with the RNA. The details of the algorithm are discussed in the following section. The rationale underlying this approach is to create an algorithm that is able to avoid calculating the ensemble of possible assembly intermediates (which we will show increases by many orders of magnitude due to the RNA) yet still be able to consider all possible intermediates during assembly without the need to neglect higher-energy, more unstable intermediates [33]. Using this approach we show that different distributions of packaging signal strengths across the genome result in significantly different assembly efficiency, illustrating their crucial roles in the assembly process.

III. THE RNA VIRUS COASSEMBLY MODEL BASED IN A GILLESPIE FRAMEWORK

In contrast to previous models that treat RNA as a homogeneous polyelectrolyte [21–25], we treat the RNA as a heterogeneous structure containing 12 binding sites as shown in Fig. 2(a), where the RNA-protein complexes between the RNA and capsid proteins have different lifetimes, depending on their location (denoted by $-6,\ldots,6$ from left to right, reflecting central symmetry in the absence of nucleotide information) in the RNA. We model capsid assembly following insights from recent experiments [15]. Initially, capsid proteins bind to and unbind from the RNA, until a nucleation event around a strong packaging signal leads to the ordered addition of coat protein according to the local-rule model described above. To accomplish this, we keep track of the binding and unbinding of coat proteins to and from different parts of the RNA and, *in addition*, monitor capsid intermediates and the different RNA configurations possible for each, as shown in Figs. 1(e) and 1(f).

In order to contrast our algorithm with previous work, we briefly discuss traditional kinetic or stochastic assembly models. In traditional kinetic models of capsid assembly, one first determines all possible capsid intermediates [2,4] (or a reduced set [33]) that can be formed during assembly and then infers all possible reactions between these intermediates. These reactions are typically of the form

$$C_i + C_j \rightleftharpoons C_k, \tag{1}$$

where C_i denote the assembly intermediates. In traditional protein-centric approaches one then derives a set of N coupled differential equations for the concentrations of the assembly intermediates,

$$\frac{dC_i}{dt} = k_b C_k - k_f C_i C_j + \cdots, \tag{2}$$

where k_f and k_b represent forward and backward rates for the reactions. These are integrated, e.g., via Runge-Kutta 45 [16], to obtain a plot of the concentrations $C_i(t)$ for each of the intermediates.

For our dodecahedral study system, we have determined the complete list of assembly intermediates for the capsid with and without RNA to illustrate the drastic increase in complexity that occurs upon inclusion of RNA-protein binding events. For each intermediate containing $n \in \{1,\ldots,12\}$ pentamers, Table I shows the number of unique intermediates for the capsid protein-only case (total of 73) and how this increases when the RNA is included (total of 85 376). For larger systems, the complexity of the number of intermediates increases dramatically. Even for a $T = 1$ structure formed from 30 dimers rather than 12 pentamers, the number of intermediates for the capsid alone exceeds 2.4×10^6, with over 35.6×10^6 reactions (both backward and forward) between them. For slightly larger capsids, such as the $T = 3$ capsid of MS2 built from 90 dimers, the total number of unique assembly intermediates (again for the protein-only case) is estimated to be in the range of 10^{12}, with $\approx 10^{15}$ possible reactions between them. Although the number of dimers has only tripled between these two scenarios, this is sufficient to make the calculation of all intermediates and reactions essentially impossible without a significant increase in computational power. Indeed, for the more complex assembly scenario in the presence of genomic RNA, the construction of the protein intermediates plus RNA layouts, followed by application of the integration step in Eq. (2), is not practicable. Traditional methods to circumvent this problem rely on intermediate pruning, i.e., keeping only the more stable intermediates in order to reduce the size and complexity. For example, Endres *et al.* [33] suggested that reducing the ensemble of intermediates considered in the model to about 15%–20% of the energetically most stable ones would be sufficient to account for capsid assembly kinetics.

We use here a method that does not require explicit construction of any intermediate list and hence allows the additional complexity of RNA binding events to be incorporated, even in large viral capsids. In particular, we use a Gillespie stochastic simulation algorithm within an object-oriented programming paradigm that allows for the simultaneous

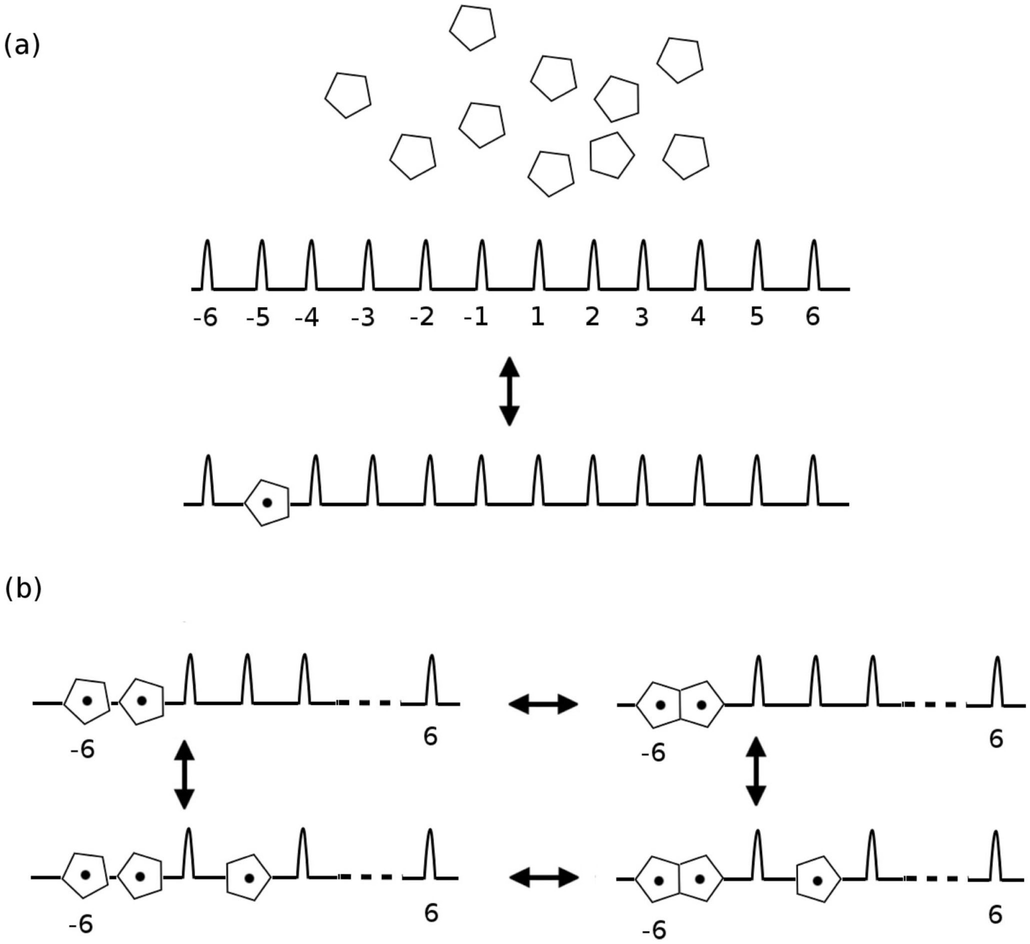

FIG. 2. Possible assembly reactions in the model. (a) The RNA is modeled as containing 12 packaging signals of varying affinity, numbered from −6 to 6. Binding can occur at any of the unoccupied sites at any time in the simulation. (b) Protein pentamers and RNAs react via two different types of events: the second-order reaction of a protein-RNA association (vertical arrows) where coat proteins bind and unbind to RNA or a first-order reaction where protein-protein interactions (horizontal arrows) are formed or broken.

modeling of thousands of individual RNAs, each containing a potentially unique distribution of packaging signals. In particular, we construct a class (see a generic pseudocode for the class object in the Appendix) that contains information about the configuration of the RNA, and partially formed capsid intermediates bound to it, for every RNA present in solution. This class only requires a small amount of computer memory to be able to store the configuration of the partially formed capsid as well as a simple description of the RNA configuration. Additionally, the class contains subroutines that can calculate the set of reactions available to the RNA and partially formed capsid, depending on its current configuration. Thus each RNA-capsid (RC) intermediate in the simulation and its possible reactions are monitored via a unique class object. This algorithm extends similar algorithms for empty capsid assembly, such as the queuing-based algorithm of Schwartz and co-workers [26,27], to the situation where RNA or other polymers will cooperatively coassemble with the capsid.

Our procedure works as follows. One first calls the class subroutine that calculates the reactions available to each RC intermediate α, where α ranges from 1 to the number of RNAs in the simulation. The subroutine calculates the set of all M_α reactions available to the RC intermediate, the reaction probability for each, and the total probability that any of these reactions will occur. The reaction probabilities $a_\mu(\alpha)$, with $\mu = 1,\ldots,M_\alpha$, for each reaction available to the RC intermediate are estimated from the forward and backward rates of the reaction and the number of capsid building blocks (in this case pentameric units) present in solution. For example, the forward part of the reaction [Eq. (1) above] has a probability $a_\mu(\alpha) = n_i n_j \kappa^1$, where n_i and n_j are the number of intermediates C_i and C_j present and κ^1 is a probabilistic rate with units of s^{-1}. For a well mixed volume V, the probabilistic rate can be estimated from the kinetic rate k_f and the volume of the system as $k_f = V\kappa^1$. The total probability that any reaction will occur $a_0(\alpha)$ is computed as the sum of all reaction probabilities $a_\mu(\alpha)$, i.e.,

$$a_0(\alpha) = \sum_{\mu=1}^{M_\alpha} a_\mu(\alpha). \tag{3}$$

TABLE I. Number of intermediates in the dodecahedron system with and without RNA. The number of unique intermediates for the dodecahedron with and without RNA is shown for each number of pentamers in the partially formed capsid (1–12). (a) The RNA organizations are the number of different RNA arrangements for all of the unique protein intermediates containing a given number of pentamers [column (d)]. Examples of two RNA organizations for $n = 4$ are shown in Figs. 1(e) and 1(f). (b) The number of RNA binding configurations is given by $\sum_{M=0,N_{\rm PS}} N_{\rm PS}!/(N_{\rm PS} - M)!M!$ and is the combinatorial number of different RNA-CP complexes that can occur, given the number of available packaging signals $N_{\rm PS}$. For the case when $n = 10$, there are only two PSs remaining in the RNA that are not in complex with a capsid with four possible binding configurations, one where both PSs are bound to CP, one where both are free, and two where only one PS is CP-free. (c) The number of unique RNA-capsid intermediates considered in our algorithm when RNA is present. This is the product of columns (a) and (b). (d) The number of intermediates in the RNA-free case. This column is equivalent to the number of dodecahedral intermediates modeled in Zlotnick's approach [2,3].

Pentamers in capsid intermediate	RNA organizations (a)	RNA binding configurations (b)	RNA-capsid intermediates (c)	Protein intermediates (d)
1	1	4096	4096	1
2	1	1024	1024	1
3	4	512	2048	2
4	14	256	3584	5
5	46	128	5888	9
6	142	64	9088	20
7	396	32	12672	13
8	948	16	15168	12
9	1832	8	14656	5
10	2672	4	10688	3
11	2600	2	5200	1
12	1264	1	1264	1
Total			85376	73

After calculation of the possible reactions and their probabilities, our procedure operates by randomly selecting one of the $M = \sum_\alpha M_\alpha$ reactions to fire at a random time in the future according to the probability function [32]

$$P(\tau,\mu,\alpha) = a_\mu(\alpha)e^{-\bar{a}_0\tau}, \tag{4}$$

where $\bar{a}_0 = \sum_\alpha a_0(\alpha)$ is the sum of total reaction probabilities for each RC intermediate. The probability in Eq. (4) corresponds to the probability that one of the M_α individual reactions $\mu = 1,\ldots,M_\alpha$ available to the RNA-capsid intermediate α will occur within a time τ from the current time. After the reaction μ and the RC intermediate α is selected for firing, the configuration of the RC intermediate α is changed appropriately, the time is updated by τ, and the procedure repeats.

There are two types of basic reactions available to any given RC intermediate, depending on its configuration. First, capsid protein can bind to or unbind from any of the PSs in the RNA with on and off rates of $\kappa_R^1(i)$ and $\kappa_R^2(i)$, respectively, as depicted by the vertical arrows in Fig. 2(b). Note that each PS i, with $i \in \{\pm1,\ldots,\pm6\}$, is allowed to have a unique rate in the simulation. Second, CP in complex with RNA can bind to the growing capsid shell via protein-protein interactions with on and off rates of κ_P^1 and κ_P^2, respectively, as depicted by the horizontal arrows in Fig. 2(b). The RNA-CP complexes are confined to binding at the edges adjacent to the last RNA-CP complex added to the growing capsid shell, consistent with the local-rule model discussed above [see Figs. 1(d)–1(f)]. As a result, the path followed by the RNA as it makes contact with the protein shell forms a Hamiltonian path in the fully assembled capsid [5,30,34], where the graph for the Hamiltonian path can be represented by a (hypothetical) polyhedron with vertices at the binding sites.

In the example discussed here, the RNA-protein contacts are modeled as being located at the fivefold axes of icosahedral symmetry, defining an icosahedron as the overall (icosahedrally averaged) layout of the RNA in contact with capsid. This is consistent with experimental evidence for a number of RNA viruses, for which icosahedrally averaged cryo-EM has revealed a polyhedral shell organization for the genomic RNA in contact with capsid protein [31,35–37]. Hamiltonian paths label the different possibilities in which these averaged densities can be realized by the packaged genomic RNA. Different RNA organizations in our dodecahedron model hence correspond to the different possible Hamiltonian paths on the icosahedron. There are 1264 such Hamiltonian paths for the icosahedron, each labeling a different assembly outcome. Our stochastic approach can be used to predict the relative multiplicity of occurrence of these different configurations, hence providing predictive information on the tertiary structure of the RNA in contact with the capsid shell in the fully assembled particles.

IV. RESULTS

We have used the approach outlined above to analyze the assembly of the 12 pentamers in the dodecahedral study system both with and without an RNA containing 12 packaging signals. We consider an ensemble of 3000 RNAs and 36 000 pentamers that satisfies the stoichiometry of the dodecahedral capsid. In the RNA coassembly simulations, RNAs recruit pentamers and assemble capsids according to the reactions illustrated in Fig. 2 and RNA-protein contacts are assumed to take place at the center of each of the pentamers in the dodecahedral capsid.

A. Choice of model parameters

We use the following formulas to determine on and off rates in our simulations:

$$\frac{\kappa_P^1}{\kappa_P^2} = Ce^{-\beta\Delta G_P}, \tag{5}$$

$$\begin{aligned} \frac{\kappa_R^1(i)}{\kappa_R^2(i)} &= Ce^{-\beta\Delta G_R(i)}, \\ \frac{\kappa_P^1}{\kappa_P^2} &= e^{-\beta\Delta G_P}, \end{aligned} \tag{6}$$

where C in Eqs. (5) and (6) is a dimensionless factor. Equation (5) is used in the RNA-free case, where protein-protein interactions are of second order, while Eq. (6) is used in the RNA coassembly model, where RNA must recruit CP in a second-order reaction before protein-protein associations occur via a first-order reaction. In the RNA-free case, we use the diffusion-limited kinetic rate $k_f = 10^6$ M^{-1} s^{-1}, the value used by Zlotnick [4] for protein capsid assembly, to approximate the on rate using the relation $\kappa_P^1 = 10^6/V = 0.0024$ s^{-1}, where $V = 0.7$ μm^3 is an estimate of the volume of a small bacterial cell. The factor C in Eq. (5) is adjusted such that favorable capsid assembly occurs around $\Delta G_P = -2$ kcal M^{-1}, consistent with estimates of protein-protein association energies in capsids. In the RNA coassembly model, the RNA must first recruit CP via a second-order PS binding event. We use the same on rate as in the protein-only situation, i.e., $\kappa_R^1(i) = 0.0024$ s^{-1}, and estimate the range of values for the free energy of binding possible for RNA-CP interactions, $\Delta G_R(i)$, from stopped-flow kinetic measurements of RNA stem-loop binding to coat proteins in bacteriophage MS2 [38]. These measurements determine the upper limit on RNA-CP binding, because they correspond to the strongest packaging signal in the experiment, giving a value of $\Delta G_R = -12$ kcal M^{-1}. Finally, we use on rates of $\kappa_P^1 = 100$ or 10^6 s^{-1} for the first-order protein-protein association reactions. Simulations using a single processor usually take less than a minute of computer time to reach 1000 s of simulation time.

B. The RNA-free assembly versus assembly in the presence of RNA

We start by comparing assembly kinetics in the absence and presence of RNA. Figure 3 shows plots of the percentage of capsid assembled in thermodynamic equilibrium (capsid yield, solid black curve), given the available number of pentamers, and the time (T_{90}, dashed red curve) until 90% of this yield value is reached for different choices of protein-protein association energies. The plot in Fig. 3(a) for the RNA-free case reveals a range of ΔG_P, shown delimited via dashed lines, for which assembly yield and speed are simultaneously high. *In vivo*, the need to achieve a high yield of stable capsids in a relatively short time would confine protein-protein association energies to this limited range of values.

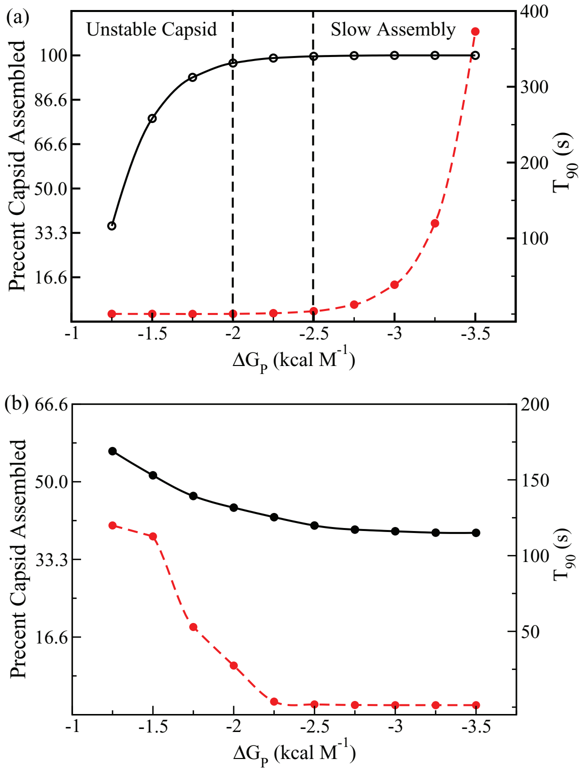

FIG. 3. (Color online) Comparison of the effects of protein association energies on assembly yield with and without RNA interactions. (a) Percentage of capsid assembled from available pentamers (solid black curve) and time to reach 90% of the maximum amount of capsid assembled (dashed red line) at thermodynamic equilibrium for a protein-only assembly scenario. Vertical lines indicate an optimal window of ΔG_P values that result in efficient assembly, i.e., high yield and speedy assembly reaction. (b) Percentage of capsid assembled in the presence of RNA for $\kappa_P^1 = 100$ s^{-1} shows that weaker protein-protein association energies can increase the yield at the expense of overall capsid stability. Data points (dots) in (a) and (b) are averaged over 100 assembly simulations. Data for the percentage of capsid assembled are taken at $t = 1000$ s, which is thermal equilibrium for all points except those lower than $|\Delta G_P| = 2.5$ kcal M^{-1} in (b), where the slow annealing occurs.

In contrast, Fig. 3(b) shows simulations in the presence of an RNA containing 12 PSs of identical CP affinity of -12 kcal M^{-1} and with the capsid nucleation site at position -4. For this scenario of a coassembly process, the percentage of capsid assembled is always low compared to the protein-only case. Moreover, it exhibits a principally different assembly behavior: Capsid yield is higher at lower values of $|\Delta G_P|$, contrary to the RNA-free case. For high values of $|\Delta G_P|$, the assembly reactions follow sigmoid kinetics and achieve equilibrium quickly ($\sim$100 s). Since we are able to track each assembling RNA via our method, we are able to assess in molecular terms the mechanism of capsid buildup in this scenario. An analysis of these data reveals that under these conditions the majority of the RNAs form partial, aberrant capsid-like structures, in which both ends of the RNA are trapped, i.e., in which both ends are surrounded by pentamers containing an RNA contact and thus cannot move to a nearest-neighbor pentamer that is unoccupied as required by the local-rule model. For low values of $|\Delta G_P|$ (<2.5 kcal M^{-1}), assembly occurs in two phases, a rapid phase of sigmoidal kinetics, followed by an extremely slow annealing phase where the partially formed aberrant structures slowly anneal to form closed capsids. Figure 3 suggests that reducing the protein-protein association energy could be a vehicle to achieve higher capsid yield. However, such an assembly scenario is likely to be inefficient due to this slow annealing phase. Moreover, the decrease in capsid yield for more negative protein-protein association energies suggests that in order for an ssRNA virus to achieve a high capsid yield, capsid stability must be sacrificed. Since this is most certainly nonviable *in vivo*, we instead explore alternative ways of increasing assembly efficiency by adjusting the PS affinities, the position of the packaging signal where nucleation starts, and the on rates. We note that the inhibitory effects of the RNA described above are, at first sight, counterintuitive. However, in biology inhibitory effects are often seen in systems where regulation of activity is important. We therefore examine in the following if the packaging signal affinities and the location of the assembly-initiating packaging signal impact on capsid yield.

C. Dependence of assembly efficiency on the position of the strong packaging signal

We explore first the dependence of assembly kinetics on the position of the high-affinity packaging signal that initiates assembly. Indeed, the position of this assembly initiation site may vary widely for different viruses. For example, the genome of bacteriophage MS2 contains a single high-affinity packaging signal TR, located near the center of the 3569-nucleotide-long genomic RNA at nucleotide 1754 [19], while for other ssRNA viruses, such as the plant virus Turnip yellow mosaic virus, the high-affinity PS is positioned towards the $5'$ end at nucleotide 18 [39]. Therefore, we explore whether capsid geometry favors a specific location of the high-affinity packaging signal that nucleates assembly. For this, we use a protein-protein association energy of $\Delta G_P = -2.5$ kcal M^{-1} and vary the position of the high-affinity signal across the 12 possible binding sites, keeping all other packaging signal affinities fixed at a weaker value of -2 kcal M^{-1}. Figure 4 summarizes the result. The percentage of capsid assembled from available pentamers and RNAs varies significantly depending on the location of the strong packaging signal, with positions -5 and -4 (and 4 and 5 by symmetry) resulting in the highest assembly yields for this choice of ΔG_P. This suggests that the position of the high-affinity PS should have an impact on assembly efficiency.

D. Dependence of assembly efficiency on packaging signal affinities

Given the impact of the position of the strong PS on assembly efficiency, we next examine how varying the binding affinities of the remaining 11 PSs further affects the assembly process. An advantage of our object-oriented algorithm is that it allows the binding affinities of all 12 packaging signals to be specified for each RNA in the simulation without adding to the complexity. In Ref. [9] we have identified multiple, lower-affinity PSs in the genome of bacteriophage MS2. In order to investigate the implications of a genome containing multiple PSs of varying affinity on capsid protein assembly kinetics, we explore the impact of different choices for the binding affinities of the remaining 11 PSs in our dodecahedron model. For an RNA containing 12 PSs that can be varied in increments of 0.1 kcal M^{-1}, there would be $120^{12} > 10^{24}$ possible packaging signal configurations, making a systematic exploration of the PS phase space unfeasible. However, even though it is not possible to explore the entire phase space of all possible combinations of PS affinities due to the complexity of the problem, it is possible to randomly sample different RNA configurations from this phase space of RNA configurations to obtain a simplified picture of how varying PS affinities affects capsid yield. For this we have randomly selected 300 RNAs from the ensemble of possible packaging signal configurations, i.e., for each RNA we randomly chose affinities for each of its 12 PSs between -0.1 and -11.9 kcal M^{-1}, fixing the strong packaging signal that nucleates assembly at position -4 at -12 kcal M^{-1}. For each of these RNAs, we have computed the capsid yield (i.e., percentage of capsid assembled at equilibrium) using a protein-protein association energy $\Delta G_P = -2.5$ kcal M^{-1} and an on rate for protein-protein association of $\kappa_P^1 = 100$ s^{-1}. The left peak in Fig. 5(a) illustrates the range of capsid yield for this ensemble of 300 RNAs given $\kappa_P^1 = 100$ s^{-1} (horizontal axis) and the probability of finding an RNA in this ensemble that assembles with this yield (vertical axis). The figure demonstrates that the few rare RNAs at the upper limit of the distribution reach a maximum capsid yield of around 66%. This suggests that for this choice of the on rates, high yields close to 100% as in the RNA-free case are highly unlikely, either because such RNAs are extremely rare and have not been sampled or because there is no distribution of PSs that will achieve a high yield without additional adjustment of other parameters.

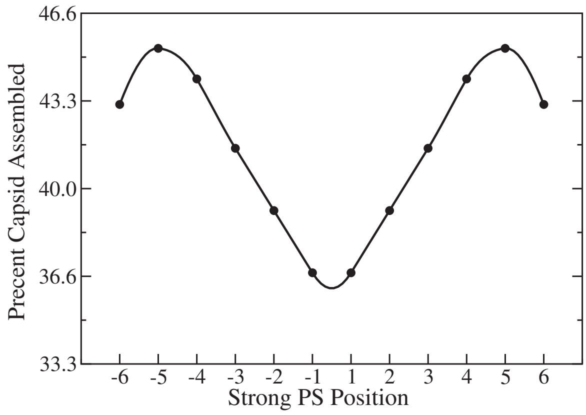

FIG. 4. Assembly efficiency depends on the location of the high-affinity packaging signal. The percentage of capsid assembled for $\kappa_P^1 = 100$ s^{-1} from available RNA and pentamers is illustrated for different locations of the strong packaging signal, i.e., the assembly initiation point, in the RNA. The highest yield of completed capsid is achieved when the strong packaging signal is placed at positions -4 and -5 (or the symmetric positions 4 and 5). Data points (dots) are averaged over 100 simulations and splined to produce the curve shown.

We therefore explore next the effect of varying the on rates. Estimates for the on rate of RNA-CP binding $\kappa_R^1(i)$ are

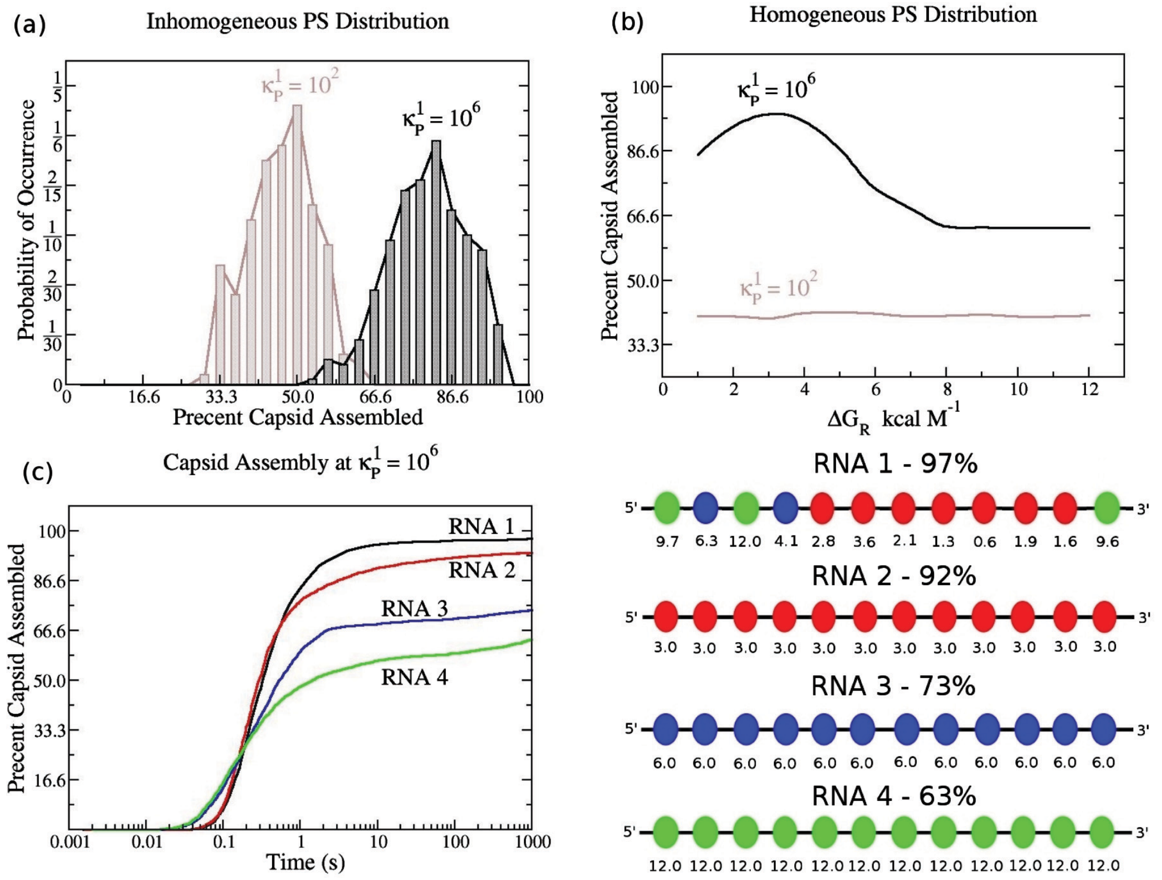

FIG. 5. (Color online) Exploring the effect of κ_P^1 and packaging signal distribution on capsid assembly. (a) Distribution of capsid yields for an ensemble of 300 RNAs with a random distribution of packaging signals. The distribution shows the probability of finding an RNA in the ensemble of 300 that achieves a given yield of capsid. The leftmost peak is computed for $\kappa_P^1 = 100$ s^{-1}, while the rightmost peak is for $\kappa_P^1 = 10^6$ s^{-1}, illustrating the effect of the parameter on capsid assembly. (b) Capsid yield for the case of a homogeneous RNA containing 12 PSs with identical affinities. For $\kappa_P^1 = 100$ s^{-1}, the yield of the capsid is roughly independent of the PS affinities, while for $\kappa_P^1 = 10^6$ s^{-1} there exists an optimal value of PS affinity at -3.0 kcal M^{-1} in which capsid yield achieves a maximum of about 92%. (c) Percentage of capsid assembled for four RNAs at values of $\kappa_P^1 = 10^6$ s^{-1} and $\Delta G_P = -2.5$ kcal M^{-1}. The RNA1 is a high-yield RNA chosen from the ensemble of 300 RNAs, while RNA2–RNA4 contain a uniform distribution of PS affinities. Although RNA1 and RNA2 have similar yields, RNA1 is more efficient at the coassembly scenario when compared to the homogeneous RNAs. Statistical data for the assembly curves were collected over 100 assembly simulations.

available from experiment [38] and we therefore keep these values fixed. Experimental estimates for κ_P^1, however, are unknown. Here κ_P^1 corresponds to the rate at which neighboring proteins, once bound to the RNA at the neighboring PS, will associate. Once two proteins are in these positions, they are already in close proximity and the rate will mainly correspond to the refolding of the small RNA portion between these neighboring PS, which is typically of the order of microseconds. Therefore, a rate of $\kappa_P^1 = 10^6$ is more realistic for RNA viruses. Results for the parameter of $\kappa_P^1 = 100$ illustrate how the packaging of materials more rigid than ssRNA would affect the assembly outcome. The second, shifted peak in Fig. 5(a) shows the outcome for the same ensemble of 300 RNAs, now assembled with an on rate κ_P^1 of 10^6 s^{-1} (keeping the protein-protein association energy fixed at $\Delta G_P = -2.5$ kcal M^{-1}). For this value of κ_P^1, the space of possible RNA configurations, i.e., of possible distributions of PS affinities, now contains a few rare RNAs that achieve a capsid yield similar to that of the RNA-free case. Similarly, as shown in Fig. 5(b), the capsid yield for homogeneous RNAs containing 12 identical PSs increases with κ_P^1. Interestingly, for higher values of κ_P^1 (e.g., 10^6 s^{-1}) capsid yield is dependent on the RNA-CP affinity, while at lower rates (e.g., $\kappa_P^1 = 100$ s^{-1}) affinity has no significant impact. In particular, when $\kappa_P^1 = 10^6$ s^{-1}, a narrow peak emerges around a weak RNA-CP affinity of $\Delta G_R(i) = -3.0$ kcal M^{-1} at which capsid yield is maximal at around 92%. This result is consistent with Brownian dynamics simulations of capsid assembly around homogeneous RNAs [22], which showed that capsid yield varies according to RNA-CP affinity, with weak interactions being important to ensure a high yield of capsid.

To contrast the difference between capsid assembly around a homogeneous RNA containing 12 identical PSs and an inhomogeneous RNA containing 12 different PSs, Fig. 5(c) shows a high-yield RNA with an inhomogeneous PS affinity distribution from the ensemble in Fig. 5(a) (RNA1) in comparison with three homogeneous RNAs containing 12 identical PSs for the case of $\kappa_P^1 = 10^6$ s^{-1} and $\Delta G_P = -2.5$ kcal M^{-1}. When averaged over 100 simulations, RNA1 assembles 2913 ± 9 viruses out of 3000 RNAs, or 97.1% ± 0.3%, while the best homogeneous RNA [containing 12 weak PSs with binding affinity of -3.0 kcal M^{-1} each (RNA2)] assembles 2776 ± 12 viruses, or 92.5% ± 0.4%. The other two homogeneous RNAs [containing 12 medium affinity PSs of -6.0 kcal M^{-1} each (RNA3) and 12 strong affinity PSs of -12.0 kcal M^{-1} each (RNA4)] assemble significantly less capsid in the same amount of time. Interestingly RNA1, which outperforms all the others, contains two high-affinity packaging signals next to the strongest packaging signal at position -4, suggesting that a cluster of higher-affinity PSs around the strong PS that initiates assembly could aid nucleation. Indeed, a similar situation was observed in the genome of bacteriophage MS2, where two additional high-affinity PSs are located a few nucleotides away from TR [9]. As can be seen from Figs. 5(a) and 5(c), a variation in packaging signal affinity can have a dramatic effect on capsid yield and the PS affinity distribution has a regulatory effect on assembly efficiency.

E. Tertiary structure prediction of RNA in contact with the capsid

The PS affinity distribution moreover determines the assembly pathways because it impacts on the position at which protein subunits are recruited to a growing capsid intermediate. Since different assembly pathways result in different organizations of the packaged RNA genome in the fully assembled viruses, our approach provides information on how the genome is likely to be organized in proximity to the capsid surface. In particular, it predicts how the RNA molecule fits asymmetrically into the RNA density in contact with capsid proteins observed in cryo-EM experiments [37].

We illustrate this here for the four cases (RNA1–RNA4) shown in Fig. 5(c). In the fully assembled capsids, the RNA connects the centers of the pentagonal faces in such a way that the RNA forms a Hamiltonian path on the inscribed icosahedron (cf. Fig. 1), i.e., it forms a connected path between the icosahedral vertices, visiting each pentagon precisely once as it does so. There are 1264 possible such paths and we investigate here the distribution of these in the assembly of protein around RNA1–RNA4. For this, we simulate the assembly of 3000 copies of each of these RNAs, repeating this process 100 times in order to generate an ensemble of roughly 200 000–300 000 fully assembled particles (note that the number is smaller than 300 000 as the yield is not 100%). These particles are then analyzed to determine the Hamiltonian path organization of their packaged genomes, i.e., the asymmetric organization of the RNA in contact with capsid protein.

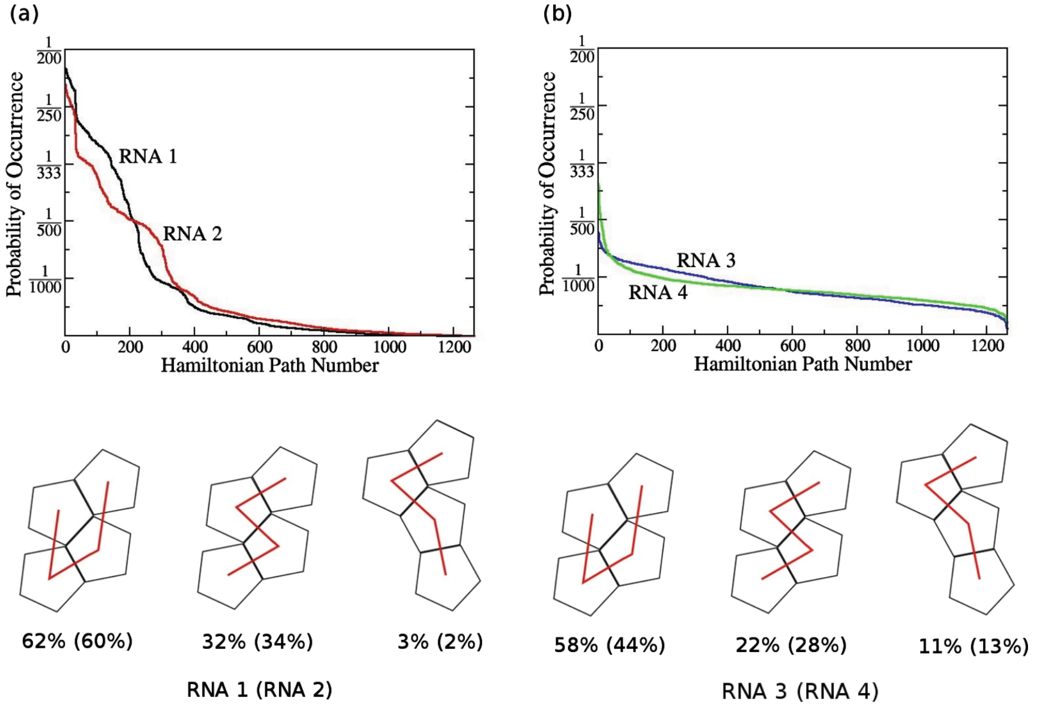

FIG. 6. (Color online) Examining the Hamiltonian path organizations of different RNAs. (a) Probability that an RNA in a capsid will be observed in one of the 1264 Hamiltonian paths. The probabilities are calculated using roughly 300 000 assembled capsids containing RNA1, a high-yield scenario with an inhomogeneous PS affinity distribution, and 300 000 assembled capsids containing RNA2, a high-yield scenario with a homogeneous PS affinity distribution. For both RNAs, there is a clear bias towards a subset of paths. Roughly 250 paths account for 75% of particles containing RNA1, while about 300 paths account for 75% of particles containing RNA2. The geometric origin of this bias is shown below the plot where 94% of RNAs pass through the two most stable intermediates with four pentamers. (b) Probability that an assembled capsid containing the low-yield homogeneous RNA3 or RNA4 will be observed in one of the 1264 Hamiltonian paths. The probability for each path is roughly 1 in a 1000, illustrating that each path is roughly equally likely. In contrast to (a) only 80% of the RNAs will pass through the two most stable intermediates with four pentamers, showing a propensity for the higher-yield RNA1 and RNA2 to bias assembly along these more stable intermediates.

Figure 6(a) illustrates the probability of occurrence of the 1264 different possible Hamiltonian paths for assembly scenarios with the high-yield inhomogeneous (RNA1) and homogeneous (RNA2) RNAs. Interestingly, there is a strong bias towards only a subset of the 1264 possible paths, with about 250 paths occurring in about 75% of particles in the case of RNA1 and about 300 paths in about 75% of particles for RNA2. By contrast, Fig. 6(b) shows that there is no significant preference for specific organizations of the packaged genomes in the lower yield cases of RNA3 and RNA4, for which all Hamiltonian paths occur with roughly equal probability (i.e., each path has a probability of occurrence of about 1 in a 1000). This suggests that there are principally different assembly behaviors for higher- and lower-yield scenarios, independent of whether the distribution of packaging signal affinities is inhomogeneous or homogeneous. The geometric reason for this difference is demonstrated in Figs. 6(a) and 6(b) in the bottom row, which shows the most abundant on-pathway assembly intermediates with four pentamers in the two scenarios. For RNA1 and RNA2, the two most stable configurations (by number of protein-protein bonds) make up 94% of all intermediates containing four pentamers, in contrast with the lower-yield cases shown in Fig. 6(b), where these two intermediates make up only about 80% of the total. Moreover, in the latter case the third intermediate, with significant probability of occurrence, has only four, as opposed to five, interpentamer bonds and is hence less favorable from a protein-protein interaction point of view. These examples suggest that RNAs that assemble with high fidelity are biased towards energetically more favorable geometries at the protein level.

Finally, we note that this selection of specific assembly pathways and genome organizations in the high-yield cases is consistent with previous experimental and theoretical results. In particular, a prediction of the contact map between capsid protein and RNA in the tertiary structure of the packaged genomes of bacteriophages MS2 and GA [9] shows a bias towards a defined conformation of the packaged genome in proximity to the capsid. Moreover, both cryo-EM reconstructions of MS2 [30,37] and an asymmetric structure determination using tomography [40] reveal that the tertiary structure of the packaged genome is highly constrained and possibly unique. Our results here suggest that this may be a consequence of selective pressures that triggered evolution of a packaging signal distribution that ensures higher capsid yield in the coassembly process, biasing the RNA to package with a more defined conformation when compared with random nonevolved RNAs.

V. CONCLUSION

The recent discovery of multiple degenerate weak packaging signals in ssRNA viruses from both bacterial and eukaryotic hosts, in addition to the small number of strong signals known to initiate capsid assembly, has opened up a new view of the assembly process in this important class of pathogens. It supports a paradigm shift away from the traditional protein-centric view of assembly to one that recognizes the full spectrum of the vital cooperative roles played by the genomic RNA. Since ssRNA viruses contain a disproportionate number of human diseases and agricultural pests affecting crops and livestock, theoretical models and experiments that contribute to a better understanding of the mechanisms underlying their formation are important for the development of novel antiviral strategies against them.

Unlike their DNA viral counterparts, which first assemble an empty capsid (procapsid) and subsequently use a terminase (packaging motor) to package their genomic material into this structure, ssRNA viruses use a sophisticated coassembly process where the RNA genome interacts with the capsid proteins during assembly. Hence assembly models that are entirely reliant on protein-protein association [2–4] are suitable only to describe the formation of empty capsids, where the cooperative roles of nucleic acids or scaffolding proteins are negligible. First attempts to incorporate capsid-RNA interactions into assembly models were based on the electrostatics between the negatively charged RNA and the positively charged capsid [21–24]. While it is possible to identify a choice of parameters in our model that yield a high amount (about 92%) of assembled capsid when the RNA is modeled as homogeneous, our results show that allowing the PSs to vary introduces a much more complex range of capsid yields that depend on the distribution of PSs affinities to coat protein. In electrostatic models, where the RNA is treated as homogeneous and the important interactions are between the negatively charged phosphodiester backbone of the RNA and the positively charged N-terminal arms of the capsid, both nonviral RNAs and viral genomes of the same length would be expected to package with equal efficiency [41]. By introducing multiple PSs with varying affinity to coat protein, our model creates a complex phase space of packaging signals that can plausibly explain the experimentally observed differences between the assembly efficiency of nonviral RNAs and native cognate genomes for a variety of viruses [15]. Indeed, this observation is consistent with our recent work on the locations of packaging signals in bacteriophage MS2, which revealed a wide distribution of packaging signal affinities with a significant proportion of signals having weak affinities [9]. The analysis presented here suggests that a distribution of widely varying packaging signal affinities to coat protein could be the result of calibrating capsid yield and assembly speed during viral evolution, conferring a selective advantage to RNAs containing an optimal distribution of packaging signals with defined affinities that had previously been overlooked. Although a difference of about 5% in yield may seem too small to confer an advantage to a virus, 5% makes a huge difference in an evolutionary context. Assuming all else equal, the 97% yield RNA1 would be assembling more viral particles than the 92% yield RNA2. Although this difference is small in the first infection cycle (RNA2 producing, e.g., 92 viruses versus RNA1 producing 97 viruses), in subsequent infection cycles this difference grows exponentially. As a result, it requires only a few infection cycles (after approximately 50 cycles, 93% of all progeny virus contain RNA1) for the vast majority of progeny viruses to contain RNA1 instead of RNA2. Given the short replication cycles of RNA viruses, these small differences in yields are relevant and important to viral evolution.

The results presented here show that ssRNA virus assembly can only be fully understood if the roles of the packaging signals are incorporated into the model. The important message of this paper is that yield and speed of assembly can be tuned by the RNA, and optimized, by multiple packaging signals with varying affinities. Moreover, our simulations reveal that nearly any generic RNA will assemble some completed capsids, explaining why assembly of noncognate RNAs can be observed in experiment. However, in the parameter ranges explored here, these scenarios generally result in lower yields than inhomogeneous packaging signal distributions. The mechanisms underlying the assembly of ssRNA can hence only be fully understood if the crucial roles of the packaging signals are appreciated.

The existence of larger classes of packaging signals with weaker capsid protein affinities and their potential roles during capsid assembly have long been overlooked. Searching for these packaging signals has largely been unsuccessful, as they share only short common sequence preference motifs and secondary structures rather than defined nucleotide sequences. An approach combining SELEX experiments, which provide information on RNA sequences with affinity to capsid protein, with functional variation experiments and a new bioinformatics approach designed to search for common structural motifs has identified such distributions of weaker packaging signals in a number of viruses [8,9]. The analysis presented here provides an explanation for their occurrence, demonstrating the importance of multiple weak packaging signals for yield and speed of assembly. This result paves the way for *in silico* experiments that analyze selective pressures, such as the need for efficient capsid assembly, on the evolution of viral RNA genomes. First results along these lines, studying how the evolution process may result in tuning of the packaging signal affinities across a viral genome, are under way via a genetic algorithm that alters packaging signals according to a fitness function defined in terms of assembly efficiency. Such studies will provide new insights into the constraints on the evolution of RNA viruses, in addition to the requirement of coding for gene products, which may ultimately open up the possibility of predicting the outcomes of RNA virus evolution.

ACKNOWLEDGMENT

E.C.D. and R.T. thank the Leverhulme Trust for funding via a Research Leadership award.

APPENDIX

We present here a brief description and pseudocode for the class used in our object-oriented approach. The basic class construct is given by the following:

```
Class RC_Intermediate

INTEGER protein(12)
INTEGER rna(12)

REAL psaffinity(12)

CONTAINS

SUBROUTINE getreactions

SUBROUTINE firereactions

End Class.
```

The integer array protein ranges from one to the number of capsid building blocks, here 12 pentamers. This array is a bookkeeping device that keeps track of which pentamers are present in the protein layer. The array `rna` stores information on which RNA-protein binding sites are in complex with coat protein and whether the RNA-CP complex on that site is associated with, or detached from, the growing capsid. Finally, the array `psaffinity` contains the on and off rates for capsid proteins to associate with each RNA site. This feature allows us to model RNAs with different packaging signal affinities within a competitive assembly scenario, allowing the possibility to evolve RNAs using assembly efficiency as a fitness function. The function `getreactions` uses the information contained in the arrays `protein` and `rna` to compute the possible reactions available to the partially formed capsid, while `firereaction` chooses one of these reactions to fire according to the probability functions in Eq. (4).

[1] C. Li, J. C. Wang, W. M. Taylor, and A. Zlotnick, J. Virol. **86**, 13062 (2012).

[2] A. Zlotnick, J. Mol. Biol. **241**, 59 (1994).

[3] A. Zlotnick, R. Aldricha, J. M. Johnsona, P. Ceresa, and M. J. Young, Virology **277**, 450 (2000).

[4] D. Endres and A. Zlotnick, Biophys. J. **83**, 1217 (2002).

[5] P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans, and A. E. Ashcroft, J. Mol. Biol. **369**, 541 (2007).

[6] G. Basnak, V. L. Morton, O. Rolfsson, N. J. Stonehouse, A. E. Ashcroft, and P. G. Stockley, J. Mol. Biol. **395**, 924 (2010).

[7] E. C. Dykeman, P. G. Stockley, and R. Twarock, J. Mol. Biol. **395**, 916 (2010).

[8] D. H. Bunka, S. W. Lane, C. l. Lane, E. C. Dykeman, R. J. Ford, A. M. Barker, R. Twarock, S. E. Phillips, and P. G. Stockley, J. Mol. Biol. **413**, 51 (2011).

[9] E. C. Dykeman, P. G. Stockley, and R. Twarock, J. Mol. Biol. (unpublished).

[10] R. Golmohammadi, K. Valegard, K. Fridborg, and L. Liljas, J. Mol. Biol. **234**, 620 (1993).

[11] O. Rolfsson, K. Toropova, N. A. Ranson, and P. G. Stockley, J. Mol. Biol. **401**, 309 (2010).

[12] D. Beckett and O. C. Uhlenbeck, J. Mol. Biol. **204**, 927 (1988).

[13] D. Beckett, H. N. Wu, and O. C. Uhlenbeck, J. Mol. Biol. **204**, 939 (1988).

[14] K. Valegard, L. Lilias, K. Fridborg, and T. Unge, Nature (London) **345**, 36 (1990).

[15] A. Borodavka, R. Tuma, and P. G. Stockley, Proc. Natl. Acad. Sci. USA **109**, 15769 (2012).

[16] V. Morton, E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, and P. G. Stockley, J. Mol. Biol. **401**, 298 (2010).

[17] S. W. Lane, C. A. Dennis, C. L. Lane, C. H. Trinh, P. J. Rizkallah, P. G. Stockley, and S. E. V. Phillips, J. Mol. Biol. **413**, 41 (2011).

[18] R. J. Ford, A. M. Barker, S. E. Bakker, R. H. Coutts, N. A. Ranson, S. E. V. Phillips, A. R. Pearson, and P. G. Stockley, J. Mol. Biol. (2013), doi: 10.1016/j.jmb.2013.01.004.

[19] G. G. Pickett and D. S. Peabody, Nucl. Acids Res. **21**, 4621 (1993).

[20] S. A. MacFarlane, M. Shanksa, J. W. Daviesa, A. Zlotnick, and G. P. Lomonossoff, Virology **183**, 405 (1991).

[21] A. Kivenson and M. F. Hagan, Biophys. J. **99**, 619 (2010).

[22] O. M. Elrad and M. F. Hagan, Phys. Biol. **7**, 045003 (2010).

[23] C. Forrey and M. Muthukumar, J. Chem. Phys. **131**, 105101 (2009).

[24] J. P. Mahalik and M. Muthukumar, J. Chem. Phys. **136**, 135101 (2012).

[25] V. A. Belyi and M. Muthukumar, Proc. Natl. Acad. Sci. USA **103**, 17174 (2006).

[26] B. Sweeney, T. Zhang, and R. Schwartz, Biophys. J. **94**, 772 (2008).

[27] F. Jamalyaria, R. Rohlfs, and R. Schwartz, J. Comp. Phys. **204**, 100 (2005).

[28] T. J. Tuthill, K. Harlos, T. S. Walter, N. J. Knowles, E. Groppelli, D. J. Rowlands, D. I. Stuart, and E. E. Fry, PLoS Pathogens **5**, e1000620 (2009).

[29] T. Lin, Z. Chen, R. Usha, C. V. Stauffacher, J. B. Dai, T. Schmidt, and J. E. Johnson, Virology **265**, 20 (1999).

[30] E. C. Dykeman, N. E. Grayson, K. Toropova, N. A. Ranson, P. G. Stockley, and R. Twarock, J. Mol. Biol. **408**, 399 (2011).

[31] K. Toropova, G. Basnak, R. Twarock, P. G. Stockley, and N. A. Ranson, J. Mol. Biol. **375**, 824 (2008).

[32] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).

[33] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, Protein Sci. **14**, 1518 (2005).

[34] W. R. Hamilton, Proc. R. Ir. Acad. **6**, 415 (1858).

[35] L. Tang, K. N. Johnson, L. A. Ball, T. Lin, M. Yeager, and J. E. Johnson, Nat. Struct. Mol. Biol. **8**, 77 (2001).

[36] S. H. van den Worm, R. I. Koning, H. J. Warmenhoven, H. K. Koerten, and J. van Duin, J. Mol. Biol. **363**, 858 (2006).

[37] K. Toropova, P. G. Stockley, and N. Ranson, J. Mol. Biol. **408**, 408 (2011).

[38] H. Lago, A. M. Parrott, T. Moss, N. J. Stonehouse, and P. G. Stockley, J. Mol. Biol. **305**, 1131 (2001).

[39] H. H. J. Bink, J. Schirawski, A.-L. Haenni, and C. W. A. Pleij, J. Virol. **77**, 7452 (2003).

[40] K. C. Dent, J. N. Barr, J. A. Hiscox, P. G. Stockley, and N. A. Ranson, Structure (unpublished).

[41] R. D. Cadena-Nava, M. Comas-Garcia, R. F. Garmann, A. L. Rao, C. M. Knobler, and W. M. Gelbart, J. Virol. **86**, 3318 (2012).