Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy

Eric C. Dykeman[a,b], Peter G. Stockley[c], and Reidun Twarock[a,b,1]

[a]York Centre for Complex Systems Analysis, University of York, York YO10 5GE, United Kingdom; [b]Departments of Mathematics and Biology, University of York, York YO10 5DD, United Kingdom; and [c]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

Edited by Peter Prevelige, University of Alabama at Birmingham, Birmingham, AL, and accepted by the Editorial Board February 27, 2014 (received for review October 16, 2013)

One of the important puzzles in virology is how viruses assemble the protein containers that package their genomes rapidly and efficiently in vivo while avoiding triggering their hosts' antiviral defenses. Viral assembly appears directed toward a relatively small subset of the vast number of all possible assembly intermediates and pathways, akin to Levinthal's paradox for the folding of polypeptide chains. Using an in silico assembly model, we demonstrate that this reduction in complexity can be understood if aspects of in vivo assembly, which have mostly been neglected in in vitro experimental and theoretical modeling assembly studies, are included in the analysis. In particular, we show that the increasing viral coat protein concentration that occurs in infected cells plays unexpected and vital roles in avoiding potential kinetic assembly traps, significantly reducing the number of assembly pathways and assembly initiation sites, and resulting in enhanced assembly efficiency and genome packaging specificity. Because capsid assembly is a vital determinant of the overall fitness of a virus in the infection process, these insights have important consequences for our understanding of how selection impacts on the evolution of viral quasispecies. These results moreover suggest strategies for optimizing the production of protein nanocontainers for drug delivery and of virus-like particles for vaccination. We demonstrate here in silico that drugs targeting the specific RNA–capsid protein contacts can delay assembly, reduce viral load, and lead to an increase of misencapsidation of cellular RNAs, hence opening up unique avenues for antiviral therapy.

The formation of a protective protein container is a vital step in most viral life cycles (1). It is a prime example of molecular self-assembly, the fundamental principle underlying the formation of protein nanocontainers that are important in virology, cell biology, and bionanotechnology (2). Whereas the molecular mechanisms of capsid formation and genome encapsulation vary across viral families, there are a number of common features that can be characterized collectively. Procapsid formation may occur via the self- or assisted assembly of coat protein subunits (CPs) and be followed by the introduction of the genomic material via a packaging motor, as seen in many double-stranded DNA viruses (3). Capsid assembly may also follow a coassembly process involving protein subunits and the viral genome, a phenomenon occurring in many single-stranded RNA (ssRNA) viruses (4). These latter are one of the largest viral families, including major pathogens (5).

The study of virus assembly in vitro has been characterized by a reductive approach, focusing on the isolation of a small number of previously disassembled components of the system and an analysis of the reactions between them. In these studies genome encapsidation is often nonspecific with noncognate RNA and even anionic polymers being encapsidated (6–10), leading to an interpretation that CPs alone have the ability to form capsids identical to those formed in vivo. However, these studies fail to address the efficiency, fidelity, and speed of assembly in vivo and also the observation that in cells genome encapsidation is highly specific (11). Additionally, they neglect the exquisite timing of events following infection, genome uncoating, gene expression, and creation of progeny genomes and eventually virions. Only late in each infection cycle is capsid assembly required and this is triggered by an increase in CP concentration.

Whereas many earlier studies have examined the assembly of protein subunits alone (12–17), we have recently emphasized the important cooperative roles played by the protein subunits interacting with the viral genomes of ssRNA viruses (4, 18). These have been shown to control the specification of the quasiequivalent conformations of the CPs, e.g., via dynamic allostery in bacteriophage MS2 (4, 19) and the assembly pathways each system follows. It appears that many RNA viruses have multiple sites within their genomes, called packaging signals (PSs) that contact CPs during assembly (18, 20, 21). The sequences/secondary structures of PSs are degenerate and they cannot be identified by sequence analysis alone. They have a range of CP affinities that have been tuned by evolution to promote productive assembly of the protein shell. We have identified such PSs in a number of ssRNA viruses and demonstrated the molecular basis of some of their effects, including in the RNA phages (4, 21) and for the plant satellite virus satellite tobacco necrosis virus (STNV) (22, 23). Single molecule assays of assembly of these viruses in vitro recapitulate the specificity of genome encapsidation seen in vivo (21), presumably because they are done at very low concentrations, approximating those in vivo, and directly demonstrate how CP–PS interactions impact on assembly as a collective. Single copy high-affinity PSs have been reported for a wide range of viruses (24, 25). Such genomes are very likely to also contain unrecognized lower affinity sites, suggesting that CP–PS interactions are more common than widely believed. Although these studies provide important insights into the mechanism of capsid assembly, the fundamental

Significance

One of the important puzzles in virology is how viruses assemble protective protein containers for their genomes rapidly and efficiently during an infection. Recent advances in the field of RNA viruses suggests that multiple specific contacts between the genomic RNA and the proteins in these containers play crucial roles in this process, but the detailed molecular mechanisms by which this occurs are largely obscure. We describe here a mathematical model of virus assembly that incorporates these contacts and other details of real virus infections. It demonstrates how such contacts act collectively to reduce the complexity of virus formation, ensuring efficient and selective packaging of the viral genomes. These competitive advantages shed new light on viral assembly and evolution and open up unique avenues for antiviral therapy.

Author contributions: E.C.D., P.G.S., and R.T. designed research; E.C.D. performed research; E.C.D. and R.T. contributed new reagents/analytic tools; E.C.D., P.G.S., and R.T. analyzed data; and E.C.D., P.G.S., and R.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.P. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: Reidun.Twarock@york.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319479111/-/DCSupplemental.

reasons for the observed assembly efficiency in vivo remain largely opaque. In particular, it is still unclear how viruses manage to overcome the limitations of the viral equivalent of a Levinthal's paradox, i.e., the selection of the assembly intermediates and pathways best suited to enhance viral load and the navigation among the vast numbers of possible pathways and potential kinetic traps in a limited amount of time. Even for a relatively small virus, e.g., bacteriophage MS2 with a capsid formed from 90 protein dimers (26), there are on the order of 10^{15} different possible intermediate species. If each of these were explored, capsid assembly would be excessively slow, or even worse, potentially result in kinetic traps where capsid proteins form stable off-pathway intermediates. This contrasts with what is observed in nature; assembly of capsids is highly efficient and robust to kinetic trapping. Previous models that exemplify the roles of CP–CP interactions and, where applicable, interactions with genomic RNAs, only partially address these issues (7–10). We show here that a vital step in resolving the viral assembly paradox lies in the timing of viral CP production, the protein "ramp," i.e., the accumulation of CP that would naturally occur in the in vivo process. In particular, we demonstrate that the protein ramp directs assembly to only a limited number of assembly intermediates and pathways, and that it also ensures packaging specificity in a cellular environment. These results suggest a paradigm shift in our approach to studying capsid formation toward in vitro and in silico assembly studies that mimic important regulatory mechanisms that occur in vivo, such as the protein ramp.

Results

The Model System. In previous modeling studies of capsid formation, it has proven fruitful to demonstrate the physical principles underlying the assembly process using the simplest possible capsid geometry, a dodecahedron formed from 12 pentagonal protein building blocks (12, 13, 27). Such a system is capable of exemplifying the assembly of capsids with more complex geometries, while still being amenable to an in depth analysis of the assembly pathways. Following this strategy, we consider in silico assembly of pentagonal building blocks into dodecahedral containers mediated by contacts with PSs in a hypothetical RNA molecule (Fig. 1). The PSs bind to the centers of the pentagons with varying affinities, and the bound CPs can then form additional CP–CP interactions, thus modeling a coassembly process during which the RNA is packaged as the capsid container forms. In this system, assembly is based on two local rules (27): (*i*) the attachment or detachment of CP units to or from the PSs present on the RNA and (*ii*) the association or disassociation of CP units bound to neighboring PSs (Fig. 1*B*). Our model defines assembly as starting via nucleation of the first CP–CP interaction on the RNA. This is consistent with known in vitro assembly processes (21, 28). So that the simulation reflects what may happen in reality, the range of affinities for these intermolecular contacts are drawn from experimental values or inferred by us and others (18, 29, 30) for the PS–CP contacts in bacteriophage MS2 and from ref. 13 for CP–CP contacts, respectively.

The exclusion of CP–CP interactions independent of RNA binding reflects experimental results showing that PS interactions with CP significantly accelerate protein–protein association. For example, in bacteriophage MS2, PS–CP interaction results in an allosteric conformational change, displacing a preexisting conformational equilibrium in the CP that lies heavily in favor of just one conformer (4, 19, 31, 32). As a result, assembly in vitro goes to completion in less than 1 h (21), as opposed to a largely incomplete assembly reaction taking many days in the absence of RNA (33, 34). Similarly, STNV CP will not assemble in vitro in the absence of RNA because PS–CP interactions are essential for overcoming electrostatic repulsions at the N termini of the CP (20, 22).

At each step of the simulation, all possible reactions between the assembly intermediates that are present are calculated and their relative probabilities of reacting determined. The reaction probabilities depend on the concentrations of the intermediates

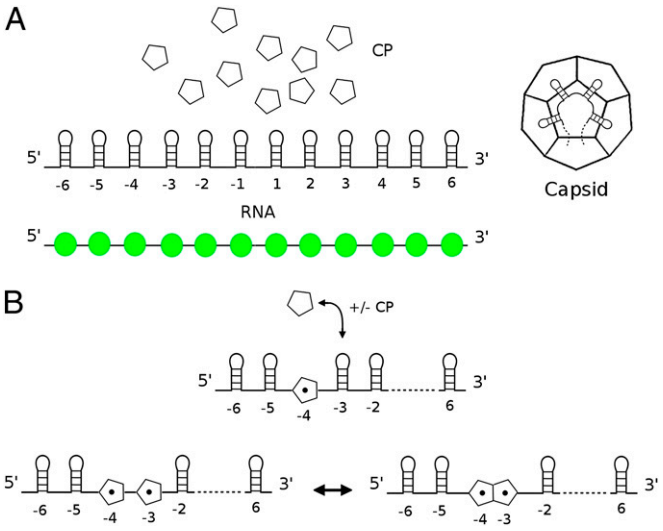

Fig. 1. The dodecahedral model of virus assembly. (*A*) The building blocks of the system are: pentagons as proxies for capsid protein (CP) and viral RNA sequences, each with 12 CP binding sites modeled as RNA stem loops (PSs), with positions labeled −6 to 6 from the 5′ to the 3′ end. Each PS can interact with CP with a different affinity. The distribution of PS affinities, here and throughout is indicated as binding free energies as a string of color-coded beads. ΔG_b values (K_d values) of $-9.5 > \Delta G_b > -12.0$ kcal/M (100 nM > K_d > 1.5 nM) are shown as green, $-5.4 > \Delta G_b > -9.5$ kcal/M (100 μM > K_d > 100 nM) as blue, and $-3.0 > \Delta G_b > -5.4$ kcal/M (6.2 mM > K_d > 100 μM) as red, respectively. (*B*) The two basic reactions of the assembly model: CP binding to RNA and subsequent formation of CP–CP contacts.

involved in the reaction and either the CP–RNA or CP–CP affinities, expressed in terms of binding free energies, as explained in *Methods*. The resulting assembly kinetics is sampled via a Gillespie-type algorithm (27, 35), i.e., at each step of the simulation the next reaction that occurs is sampled in a stochastic manner based on relative probabilities of the reactions. Time is incremented in units of seconds based on the reaction rates, which capture the average time for a single reaction to occur.

We use a Gillespie-type algorithm (35), because modeling assembly by numerically solving differential equations for the assembly kinetics is not practicable due to the complexity of the problem. The Gillespie method is scalable to real viral systems and allows us to record the assembly pathway for each RNA in the simulation. Gillespie algorithms are designed for well-mixed systems. Assembly in the presence of a linear polymer is not neccessarily so, but because PSs are located on the periphery of an otherwise compact RNA molecule and are in constant exchange with CP, we assume this to be a good approximation. We also expect local concentration effects due to cooperative binding of CPs to PSs (36) and binding accessibility of PSs due to steric effects to be negligible, based on data for MS2 (18, 21).

The Effects of the Protein Ramp on the Kinetics of Assembly. Previous work on capsid formation in silico considered the assembly of CPs in isolation, essentially mimicking an in vitro assembly scenario in which the rate of CP expression from the genomic RNAs via ribosomes is neglected. We provide here an in silico model of virus assembly in which the speed of protein subunit synthesis (the protein ramp) is included. To achieve this, we modeled a CP ramp following Eigen's example used to discuss a viral hypercycle with respect to bacteriophage Qβ (37), which was based on experimental determination of the amounts of CP, RNA templates, and infectious phage throughout the complete viral lifecycle. The experiment and Eigen's model suggest that enough CP is produced in roughly the last 5–10 min before cell lysis to assemble ~10,000–20,000 Qβ progeny. We have also shown that RNA phages assemble using multiple PSs to improve efficiency and fidelity (18, 21, 27). Using a relative low estimate for

the progeny produced per round of infection, we implement a linear rate of CP addition to 3000 hypothetical RNAs (Fig. 2), where enough CP would be titrated into the system over 5 min to complete assembly on every RNA (*Methods*). The RNAs used in this analysis are identical copies of the most/least efficient assembly template identified (RNA1/2; Fig. 2*A*), from analogous simulations of RNAs where each of the PS affinities for the RNA were randomly chosen to be a negative value above −12 kcal/M, the value of the highest-affinity PS in bacteriophage MS2 (29).

Assembly of 3000 copies of RNA1 in the presence of 36,000 CP subunits was simulated with and without the protein ramp. An intercapsomere bond strength of −3.75 kcal/M (corresponding to a K_d of 1.7 mM for a single CP–CP contact; note that by contrast the K_d for CP–RNA contacts is between 1.1 mM and 1.5 nM, i.e., between −4 and −12 kcal/M) was assumed following representative values given in the literature for other viral CP–CP contacts (12–15, 38). In the absence of the ramp (Fig. 2*B*), i.e., when all 36,000 CPs are present at the start, mimicking many in vitro reassembly assays, a number of lower-order on-pathway intermediates form initially and are then used up, but a large fraction (~40%) of the CP becomes kinetically trapped in malformed RNA–CP complexes, reducing the yield of correctly formed capsids. In sharp contrast, 100% of the RNAs get packaged in the presence of the CP ramp. The kinetics of assembly also differ between the two situations. Without a ramped protein gradient capsid assembly occurs far earlier than with it. Because assembly removes genomes that would be acting as templates in a real infection, earlier assembly in vivo is potentially counterproductive.

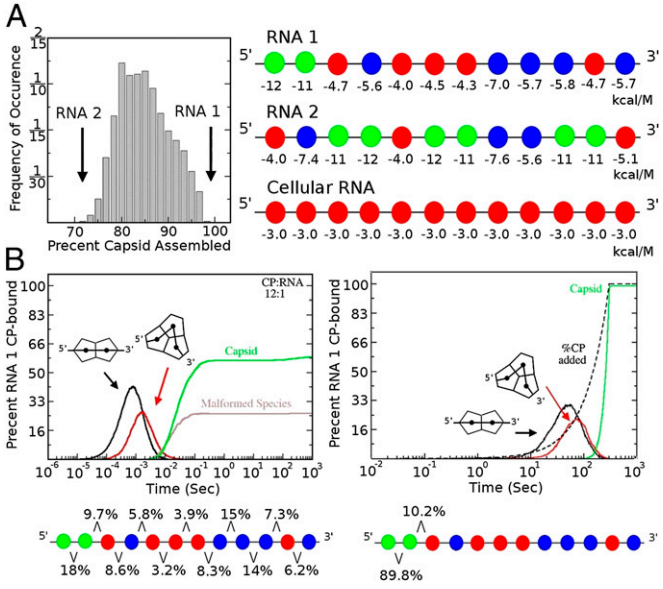

Fig. 2. The impact of the protein ramp. (*A*) A population of 3000 RNAs, each containing a random selection of PS affinities was constructed, each RNA copied 3000 times, and the percentage of RNA packaged at a CP:RNA ratio (12:1) after 1000 s calculated. The histogram shows the frequency of occurrence of RNA sequences resulting in a particular yield of capsids. The highest (99%, RNA1) and lowest (75%, RNA2) yielding RNAs in the sample were used in further simulations. A representative cellular RNA was constructed by assigning a low affinity (−3 kcal/M) to all 12 PSs. (*B*) Assembly kinetics of RNA1 determined from the average of 100 assembly simulations with 3000 RNAs at the ratio of CP:RNA (12:1). The graph shows time-dependent assembly of (*i*) assembly intermediates (protein arrangements shown as cartoons) that are on pathway and complete to capsid; (*ii*) correctly assembled capsids; and (*iii*) kinetically trapped malformed species (*Left*). In the presence of the protein ramp (CP concentration shown as a dashed line, here and throughout) all of the RNAs are successfully encapsidated. The percentage of successfully packaged RNAs that nucleate at each pair of PSs is shown below each graph. In the presence of a ramp, nucleation is confined to the two highest affinity PSs.

Moreover, the capsid assembly kinetics suggest a strong nucleation event, which is then followed by a rapid completion phase in which capsids are assembled faster than the linear production of CP. This mimics the observed assembly kinetics for Qβ infections (37) and is consistent with the results of single molecule assembly assays (21).

To investigate the reasons for this difference, we analyzed the PSs on the RNA at which nucleation takes place (Fig. 2*B*). We define nucleation as the point when formation of the initial CP–CP complex on the RNA occurs. In the absence of the ramp, nucleation may occur anywhere across the RNA. In the in vivo simulation, i.e., a CP concentration that starts low and increases, nucleation occurs principally on the adjacent, highest affinity pair of PSs (−12 and −11) (89.2% with the other pathways involving just one of these PS sites). These results show that the protein ramp facilitates coordinated nucleation, allowing assembly to proceed in a more controlled manner, as exemplified by the higher particle yield and absence of kinetic traps. To explore whether these effects were unique to RNA1, we carried out analogous simulations with the worst assembling sequence from the distribution in Fig. 2*A*, namely RNA2. The results, shown in Fig. S1, reveal formation of the same on-pathway intermediates with similar differences in assembly efficiency between the scenarios including or excluding the protein ramp. However, in this case not all RNA2s are packaged, even in the presence of the ramp. Analysis of assembly nucleation with RNA2 shows a distribution of nucleation sites both with and without ramp, albeit at multiple stronger nucleation sites in its presence. This demonstrates that RNA1 has an advantage due to its single adjacent pair of strong PSs and their role in nucleation, compared with three such sites in RNA2. These results exemplify Eigen's comment (37) that replication, which is widely used as the main criterion for fitness of viral populations in an evolutionary context, is, although important, only one of the criteria of overall performance in a viral infection. In particular, Fig. 2 shows that the interplay of the protein ramp and virus assembly has a significant impact on fitness that has not been appreciated previously.

To examine the effects of capsid geometry on these results, we also carried out a simulation for a 60mer capsid formed from 60 identical building blocks that each bind to a PS in a genome, as is the case of bacteriophage MS2. In this more complex setting, addition of the protein ramp has the effect of dramatically raising assembly efficiency (from 1 to 75%), suggesting that the 12mer model discussed here is indeed representative of more complicated viruses. We also explored the consequences of disrupting the highest affinity PSs, which control nucleation under these conditions. To do this, we reduced the affinities of these two sites to a low value (creating RNA1′; Fig. S2), hence mimicking mutagenesis experiments that impact on PS structure and affinity. Nucleation becomes uncoordinated, and assembly is incomplete (~80%), resulting in the formation of a significant fraction of misassembled species, and also in the misencapsidation of cellular RNAs.

A Reduction in Complexity. The complexity of the virus assembly process can be understood in terms of the number of possible pathways from a single protein subunit to the fully formed capsid. Because the model requires proteins to bind the RNA before attaching to other protein subunits on the growing capsid shell, assembly occurs effectively along the RNA molecule, starting at a nucleation site and then continuing toward the 5′ and 3′ ends. Exclusion of association of protein subunits that are not bound to neighboring PSs on the genomic RNA is motivated by experimental results for MS2 (21, 31). Assembly pathways can therefore be characterized by the different possible packaging arrangements of the RNA molecule within the fully formed capsid, together with specification of the initiation site. From a mathematical point of view, a superposition of all possible pathways connecting the putative RNA–CP contacts looks like a polyhedral shell, the icosahedron shown in red within the dodecahedral capsid (Fig. 3*A*). Any connected path visiting every RNA–CP contact precisely once describes a possible layout of the packaged RNA. Connected paths of this type are called

Hamiltonian paths (39), and there are 1264 such paths for the icosahedron shown in Fig. 3*A*. For larger viruses, such as bacteriophage MS2 with 90 building blocks and 60 PSs, the situation is significantly more complex with 40,678 such paths, but the principle of enumerating distinct pathways is similar. Hamiltonian paths are in a one-to-one correspondence with the organizations of the genomic RNA in the vicinity of the capsid, i.e., they specify which PS binds to which CP building block (capsomere). Because a Hamiltonian path specifies neither where assembly nucleates nor in which order the recruitment of CP subunits occurs toward the 5′ and 3′ ends, each Hamiltonian path represents an ensemble of possible assembly pathways. However, because in our simulations nucleation occurs at the two adjacent high-affinity PSs located at the 5′ end of the test RNA1, there is a unique assembly scenario for each Hamiltonian path in this case, i.e., assembly occurs linearly toward the 3′ end.

In Fig. 3*B* we plot the number of particles in the ensemble of assembled material with RNA1 with a specific Hamiltonian path for their packaged RNAs, with the *y*-axis showing the frequency of occurrence and the *x*-axis labeling the possible Hamiltonian paths, ordered from most to least frequently used. The protein ramp results in a dramatic reduction in the number of different Hamiltonian paths followed, demonstrating its role in reducing the complexity of the assembly process. As a second complexity measure, we compute for every assembly pathway, the number of intermediates in that pathway that do not correspond to the most stable intermediate (i.e., the intermediates that have less than the maximal possible number of CP–CP contacts; Fig. 3*C*). We then plot the percentage of pathways that have a given number of deviations from the most stable ones, i.e., from those on which every intermediate forms the maximum number of intersubunit bonds possible. In contrast to an assembly process in the absence of RNA, the most stable pathway is not unique because there can be different RNA configurations for any given protein configuration. Once again it appears that the protein ramp biases assembly toward the small subset of the most stable intermediates. These features provide a solution to the viral analog of Levinthal's paradox (40) for capsid assembly and illustrate how assembly in such systems is greatly aided by the natural timing of viral CP production.

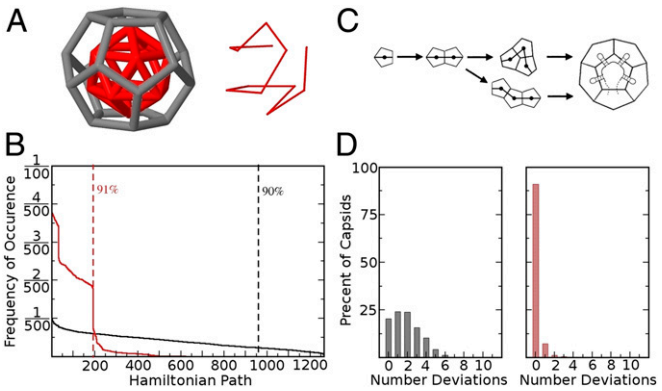

Fig. 3. Impact of the ramp on assembly pathway selection and intermediate stability. (*A*) In assembled capsids the PSs contact the midpoint of every pentagonal face of the dodecahedron and hence the RNA describes a connected path on the inscribed icosahedron (red). One such Hamiltonian path is shown on the *Right*; the icosahedron is the superposition of all possible Hamiltonian paths. (*B*) The frequency of occurrence of different Hamiltonian paths across a sample of 150,000 RNA1s determined by tracking the addition of CP on individual RNAs. The distribution in the presence of the ramp (red) shows a bias toward a small number of different paths, whereas in its absence, all possible paths occur (black). Dashed lines indicate the number of paths used by ∼90% of the 150,000 capsids. (*C*) Assembly intermediates contain more or less stable species depending on the number of CP–CP contacts. (*D*) For each assembly pathway taken the number of deviations in protein configurations from the most stable structure during assembly are shown. In the absence of the ramp (*Left*), up to six deviations can occur and their frequencies are relatively similar. In its presence (*Right*) assembly is strongly biased toward the most stable assembly intermediates.

The Protein Ramp and Selective Packaging. Assembly efficiency is one aspect of in vivo viral life cycles. It is intimately linked with packaging specificity, because cellular RNAs may act as competitors for assembling CPs, leading to subpopulations of nonviable particles and reducing the overall viral load (11). Because observations suggest that such competition is avoided, we examined whether the protein ramp plays a role in specificity. We used the same in silico assembly experiment, but introduced an element of competition between RNA1, the viral RNA, and potential cellular competitors. The latter were assigned identical, low affinities for 12 potential PSs creating RNA3 (Fig. 2). We then used an ensemble of 2000 RNA1s and differing ratios of RNA3 excess, simulating the situation in vivo. The reactions were run with just sufficient CP to fully assemble the aliquot of RNA1. The results of a simulation at a RNA1:RNA3 ratio of 1:300, in agreement with experimental estimates for the number of mRNAs in a mammalian cell (10^5 to 10^6) (41), are shown in Fig. 4. The effects of the ramp are again very spectacular. With ramped protein addition very little (<1%) of RNA3 becomes misincorporated into capsids, whereas in the absence of the ramp the competitor "cellular" RNA3s become the dominant species incorporated and this effect is robust across a wide range of concentration ratios. RNA2, the worst performing RNA from the initial random distribution in Fig. 2, is also very sensitive to the CP ramp, but less able to prevent misincorporation by RNA3 (Fig. S1*B*). In light of these results, we have examined the threshold affinities of the PSs in RNA3 required to make it an effective competitor (Fig. S1*C*). For a wide range of affinities, only the viral RNA1 is packaged, but for PS values of less than −5.5 kcal/M, RNA3 becomes a successful competitor. Note, it is unlikely that any cellular RNAs will contain an ensemble of such high-affinity sites leading to successful competition in vivo (21, 42).

The protein ramp also dramatically improves the efficiency of protein assembly in the absence of RNA (Fig. S3). This raises the interesting issue that viruses coassembling around their genomes must assemble efficiently enough to compete with a protein-only reaction. As we have shown previously, RNAs with high affinities at every PS site are poor at assembly, because they nucleate at many places and become kinetically trapped (27). Here we have shown that in the case of low affinities at every PS, e.g., as in RNA3, they are also poor substrates for encapsidation in the presence of protein ramp. It appears therefore that a combination of mixed PS affinities is the most successful strategy of encapsidation for RNAs in a coassembly process.

The simulation with ramp reflects the observed specificity of viral infections in vivo (11) and has major implications for our understanding of quasispecies evolution. The differential specificity for different members of a viral quasispecies, as exemplified here for RNA1 and RNA2, suggests that the presence of a protein ramp impacts differentially on the fitness of different members of a quasispecies (due to their distinct PS distributions). This has important consequences for the evolution of the quasispecies as a whole, which have not been appreciated previously. Finally we note that these results also provide an explanation for outcomes of in vitro assembly experiments in which the CP subunits are all added at the start of the reaction. These routinely show that it is possible to encapsidate noncognate RNAs (43), whereas our results show that in the presence of the ramp, i.e., in an in vivo scenario, this would be much less likely to occur.

A Route to Antiviral Therapy. These results suggest that drugs targeting CP interactions with the higher affinity PSs within a viral genome should have a highly disruptive effect on capsid assembly. We carried out simulations with 2000 RNA1s and the required CP concentration to form 2000 viral particles in the presence of drug molecules binding strong PSs with a K_d of 10 nM and weak PSs with a K_d of 50 μM. The protein concentration

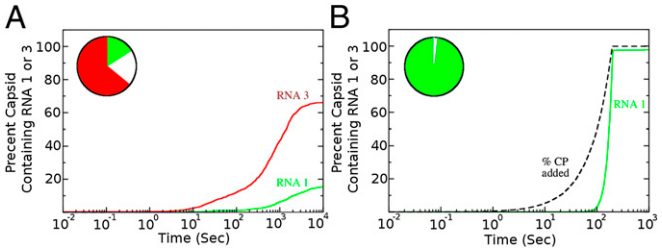

Fig. 4. Impact of the protein ramp on packaging selectivity. An in silico competition experiment between 2000 copies of RNA1 and 600,000 copies of RNA3 at a CP:RNA1 ratio (8:1) just sufficient to assemble 2000 capsids. (*A*) In the absence of the ramp, a significant fraction of RNA3 is encapsulated and the assembly yield of RNA1 is much lower due to the production of malformed species. The circular *Inset* indicates the percentages of encapsulated RNA1 (green), RNA3 (red), and malformed species (white) at the end of the simulation. (*B*) With the ramp assembly of RNA1 occurs almost to 100% with no encapsidation of RNA3, i.e., there is high packaging specificity.

is again modeled as a ramp, while the drug is administered in a single dose at the start of the simulation for different concentrations $c_D = 1$, 10, 20, and 50 μM. We observe an increasing delay in the onset of assembly with increasing drug concentration in the absence of cellular RNAs, whereas in their presence at a concentration ratio of 1:300 for RNA1:RNA3 the drug results in a significant difference in viral load (e.g., a drop of 20% at 1 mM; Fig. 5*B*) and increased amounts of misincorporated cellular RNA. Similar simulations for a number of concentration ratios of RNA1:RNA3 show: (*i*) a time delay in capsid assembly, (*ii*) a reduction in overall viral load, (*iii*) a significant population of malformed species, and (*iv*) an increase in the number of packaged cellular RNAs. In a real viral infection, malformed species or noninfectious capsids carrying cellular RNA would trigger natural antiviral defenses (immune responses and/or RNA silencing) potentially resulting in clearance of the infection. These results demonstrate that targeting PSs with drugs that inhibit their functions, especially high-affinity ones, is a promising new avenue for antiviral therapy.

Discussion

Viruses are major pathogens of people, animals, and plants, yet options for treatment and prevention are currently limited. Vaccination is the most successful medical intervention, but is likely to be restricted to a severely limited set of these pathogens. The rate at which emerging viruses develop or existing ones mutate requires urgent development of novel means to control their spread and ameliorate their effects (44, 45). Viral capsid assembly is a central process in most cases, but is as yet a largely unexploited drug target (7, 46, 47). It has strong potential because malformed virus-like particles make good immunogens that expose conserved protein epitopes, or in the case of ssRNA plant viruses, the genomic RNA would trigger host gene silencing. Previous attempts to probe the detailed molecular mechanism of viral assembly have relied on in vitro assays in which stoichiometric amounts of the assembly components have been used from the start, and many attempts to simulate assembly have used the same starting conditions. The results of such assays and simulations do not match virus production in vivo (11) and therefore fail to encompass all of the features of "real" infections. This implies the need for a more holistic view that mimics important features of the in vivo assembly process.

Here, we have simulated assembly of ssRNA viruses that coassemble their CP shells around their genomes (23, 48). Recently, we have provided compelling evidence that, at least in a number of model viruses, such assembly reactions depend on specific interactions between viral CPs and dispersed, sequence-degenerate segments of the genome (PSs) (4, 18–23), reflecting both in vitro and in vivo data for the RNA phages (49–51) and plant satellite viruses (20, 52). Given this background, the intricate interdependence of the distribution of PSs and their affinities in the viral RNA becomes an important feature to model. In addition, we have explored the consequences of having the CP concentration increase over time, mimicking what happens in a real infection. The results are dramatic. This simple effect alters the kinetics and yield of capsids produced. It does so by controlling nucleation, limiting the number of assembly pathways each particle follows, ensuring that intermediates are stable and that the virus avoids the pitfalls of the viral analog of Levinthal's paradox. Remarkably, it also has significant advantages in ensuring specificity of genome incorporation. Thus, a number of features corresponding to observed virus biology are predicted as a consequence of the PS-mediated assembly mechanism. Such predictions do not arise from more coarse-grained models of assembly pathways (7–9).

The ssRNA viruses are just one class of many viral families, but the protein ramp could potentially also have an influence on the assembly of CP subunits in isolation or that use scaffold proteins to play a role similar to the RNA, i.e., those that assemble procapsids (53). In silico experiments suggest that assembly of protein subunits in isolation is highly prone to kinetic trapping (15, 38). This has led to the suggestions that assembly involves two phases of CP–CP interactions, a slower nucleation rate, and a faster elongation rate, which takes over once a critical nucleus is formed (15). In the case of assembly of empty capsids from protein alone (Fig. S3), the protein ramp results in both the elimination of kinetic traps, as well as a reduction in the assembly pathways without the need to invoke a special nucleation rate. This result addresses the kinetic trapping puzzle observed in earlier studies of virus assembly (15, 38) by mimicry of the CP accumulation that would occur in an in vivo situation.

Our understanding of RNA biology has recently been transformed by the realization of the extent of the genome that is transcribed. Many of these transcripts interact with proteins and self-assemble into complex molecular machines controlling everything from gene expression to protein synthesis to splicing and development. Each of these complexes needs to assemble sequence specifically, and some of them are known to interact via multiple RNA motifs of varying affinity akin to viral PSs (54). Our results with this virus simulation highlight the importance of modeling the dynamic concentrations of RNAs and their protein partners in such cases. More widely, they have far reaching consequences for our understanding of one of the most fundamental processes in nature, molecular self-assembly, and also important technological implications. Protein containers play a key role in nanotechnology applications, for example as drug delivery vehicles or in the context of vaccine design. Insights on how to improve the yield of these containers and have tighter control over their assembly in vitro are therefore of particular interest. Conversely, these results open up unique avenues for antiviral therapies targeting specific groups of PSs. As shown here, such drugs would reduce viral load, i.e., the number of infectious particles formed during an infection, and at the same

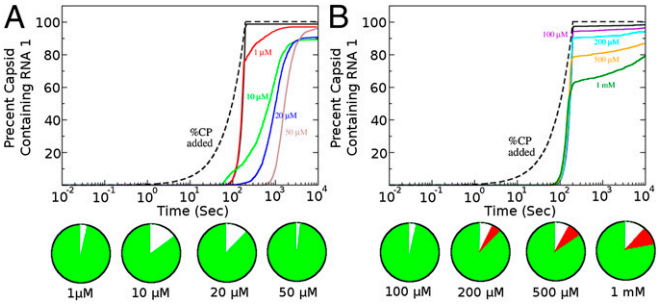

Fig. 5. Impact of PS-binding drugs on assembly efficiency. (*A*) Amounts of encapsidated viral RNAs (RNA1) in the absence of RNA3 (shown as percentages of RNA1 (green) and malformed species (white) in the presence of the ramp at differing drug concentrations. (*B*) As in *A* in the presence of cellular RNAs (RNA3, red) at a ratio RNA1:RNA3 of 1:300.

MICROBIOLOGY

time increase formation of noninfectious particles containing cellular competitor RNAs that trigger the host's antiviral defenses. Given that PSs are generic features shared across large classes of viral families, including major human pathogens (25), this is a form of drug therapy that is applicable widely and could therefore have significant impact on the health sector.

Methods

The in silico assembly experiments were performed using the Gillespie algorithm (35), which has been modified to incorporate PS–CP interactions as described in ref. 27. For the protein ramp simulations, the computational model adds CP to the simulation following a Markov chain procedure where the time interval until the next addition of a single CP is computed from $\Delta T = \ln(r^{-1})/a_0$, with r a randomly chosen number between 0 and 1. For the dodecahedral model, we use a CP synthesis rate of $a_0 = 120$ CP per second. Reactions between CP and RNA are then allowed to take place until the next CP is added, which mimics the gradual synthesis of protein subunits characteristic of in vivo assembly. For the competition simulations, 2000 RNA1 (or RNA2) were combined with 600 k RNA3 for a total of 602 k RNAs. Enough CP to assemble the aliquot of RNA1(2) was either ramped in following the above procedure or added at the beginning of the simulation. Further details about the general Gillespie procedure and specific rate choices for CP–CP and CP–PS interactions can be found in *SI Methods* and in ref. 27.

ACKNOWLEDGMENTS. We thank UK's Engineering and Physical Sciences Research Council (EP/K028286/1 and EP/K027689/1), the Biotechnology and Biological Sciences Research Council (BB/J00667X/1), and the University of York for funding.

1. Rossmann MG, Johnson JE (1989) Icosahedral RNA virus structure. *Annu Rev Biochem* 58:533–573.
2. Steinmetz NF, et al. (2009) Buckyballs meet viral nanoparticles: Candidates for biomedicine. *J Am Chem Soc* 131(47):17093–17095.
3. Sun S, et al. (2012) Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc Natl Acad Sci USA* 109(3): 817–822.
4. Stockley PG, et al. (2007) A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol* 369(2):541–552.
5. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* 60:51–67.
6. Comas-Garcia M, Cadena-Nava RD, Rao AL, Knobler CM, Gelbart WM (2012) In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *J Virol* 86(22):12271–12282.
7. Kivenson A, Hagan MF (2010) Mechanisms of capsid assembly around a polymer. *Biophys J* 99(2):619–628.
8. Elrad OM, Hagan MF (2010) Encapsulation of a polymer by an icosahedral virus. *Phys Biol* 7(4):045003.
9. Forrey C, Muthukumar M (2009) Electrostatics of capsid-induced viral RNA organization. *J Chem Phys* 131:105101.
10. Belyi VA, Muthukumar M (2006) Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci USA* 103(46):17174–17178.
11. Routh A, Domitrovic T, Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci USA* 109(6): 1907–1912.
12. Li C, Wang JC, Taylor MW, Zlotnick A (2012) In vitro assembly of an empty picornavirus capsid follows a dodecahedral path. *J Virol* 86(23):13062–13069.
13. Zlotnick A (1994) To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J Mol Biol* 241(1):59–67.
14. Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ (2000) Mechanism of capsid assembly for an icosahedral plant virus. *Virology* 277(2):450–456.
15. Endres D, Zlotnick A (2002) Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* 83(2):1217–1230.
16. Sweeney B, Zhang T, Schwartz R (2008) Exploring the parameter space of complex self-assembly through virus capsid models. *Biophys J* 94(3):772–783.
17. Jamalyaria F, Rohlfs R, Schwartz R (2005) Queue-based method for efficient simulation of biological self-assembly systems. *J Comput Phys* 204:100–120.
18. Dykeman EC, Stockley PG, Twarock R (2013) Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J Mol Biol* 425(17):3235–3249.
19. Dykeman EC, Stockley PG, Twarock R (2010) Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J Mol Biol* 395(5):916–923.
20. Bunka DH, et al. (2011) Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: Implications for the assembly of a T=1 capsid. *J Mol Biol* 413(1):51–65.
21. Borodavka A, Tuma R, Stockley PG (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc Natl Acad Sci USA* 109(39):15769–15774.
22. Ford RJ, et al. (2013) Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J Mol Biol* 425(6):1050–1064.
23. Stockley PG, Ranson NA, Twarock R (2013) A new paradigm for the roles of the genome in ssRNA viruses. *Future Virology* 8:531–543.
24. Qu F, Morris TJ (1997) Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J Virol* 71(2):1428–1435.
25. Kim DY, Firth AE, Atasheva S, Frolova EI, Frolov I (2011) Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *J Virol* 85(16):8022–8036.
26. Golmohammadi R, Valegård K, Fridborg K, Liljas L (1993) The refined structure of bacteriophage MS2 at 2.8 A resolution. *J Mol Biol* 234(3):620–639.
27. Dykeman EC, Stockley PG, Twarock R (2013) Building a viral capsid in the presence of genomic RNA. *Phys Rev E Stat Nonlin Soft Matter Phys* 87(2):022717.
28. Rolfsson O, Toropova K, Ranson NA, Stockley PG (2010) Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. *J Mol Biol* 401(2):309–322.
29. Carey J, Uhlenbeck OC (1983) Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry* 22(11):2610–2615.
30. Lago H, Parrott AM, Moss T, Stonehouse NJ, Stockley PG (2001) Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: Identification of a protein conformer unable to bind RNA. *J Mol Biol* 305(5):1131–1144.
31. Morton VL, et al. (2010) The impact of viral RNA on assembly pathway selection. *J Mol Biol* 401(2):298–308.
32. Knapman TW, Morton VL, Stonehouse NJ, Stockley PG, Ashcroft AE (2010) Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Commun Mass Spectrom* 24(20):3033–3042.
33. Beckett D, Wu HN, Uhlenbeck OC (1988) Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J Mol Biol* 204(4):939–947.
34. Beckett D, Uhlenbeck OC (1988) Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J Mol Biol* 204(4):927–938.
35. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
36. McGhee JD, von Hippel PH (1974) Theoretical aspects of DNA-protein interactions: Cooperative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J Mol Biol* 86(2):469–489.
37. Eigen M, Biebricher CK, Gebinoga M, Gardiner WC (1991) The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry* 30(46):11005–11018.
38. Endres D, Miyahara M, Moisant P, Zlotnick A (2005) Designing two self-assembly mechanisms into one viral capsid protein. *Protein Sci* 14:1518–1525.
39. Hamilton WR (1858) An account of the Icosian calculus. *Proc R Ir Acad* 6:415–416.
40. Levinthal C (1969) How to fold graciously. *Mossbauer Spect Biol Syst* 22–24.
41. Islam S, et al. (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11(2):163–166.
42. Borodavka A, Tuma R, Stockley PG (2013) A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol* 10(4):481–489.
43. Bancroft JB (1970) The self-assembly of spherical plant viruses. *Adv Virus Res* 16: 99–134.
44. Drake JW, Holland JJ (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci USA* 96(24):13910–13913.
45. Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 287(5452):443–449.
46. Zlotnick A, Porterfield JZ, Wang JC (2013) To build a virus on a nucleic acid substrate. *Biophys J* 104(7):1595–1604.
47. Hagan MF, Elrad OM, Jack RL (2011) Mechanisms of kinetic trapping in self-assembly and phase transformation. *J Chem Phys* 135(10):104115.
48. Stockley PG, et al. (2013) Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J Biol Phys* 39(2):277–287.
49. Ling CM, Hung PP, Overby LR (1970) Independent assembly of Qbeta and MS2 phages in doubly infected Escherichia coli. *Virology* 40(4):920–929.
50. Peabody DS (1997) Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2. *Mol Gen Genet* 254(4):358–364.
51. Dent KC, Barr JN, Hiscox JA, Stockley PG, Ranson NA (2013) The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* 21(7):1225–1234.
52. Larson SB, McPherson A (2001) Satellite tobacco mosaic virus RNA: Structure and implications for assembly. *Curr Opin Struct Biol* 11(1):59–65.
53. Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. *Virol J* 7:356.
54. Valley CT, et al. (2012) Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc Natl Acad Sci USA* 109(16): 6054–6059.

Supporting Information

Dykeman et al. 10.1073/pnas.1319479111

SI Methods

Our capsid assembly simulations conform to the following basic algorithm. At each discrete step in the simulation we compute two quantities: (*i*) a reaction to "fire," which is then followed by an update of the quantities of the intermediates involved in the reaction, and (*ii*) the time increment τ from the current time at which this reaction occurs (see below). This procedure is then repeated until the simulation time reaches t = 10,000 s.

To identify a reaction to fire, we first determine the set of all reactions that are possible based on the assembly protocol shown in Fig. 1*A*. We then assign to each of the list of possible reactions a reaction probability P per unit time, and the probability that it will be the next reaction to occur in the time increment dt is hence $P \cdot dt$. P depends on the number N_S of each species S involved in the reaction and its kinetic rate, i.e., for a reaction of the form A + B → C with kinetic rate K_f, the reaction probability per unit time would be $P = N_a \cdot N_b \cdot K_f / V\ s^{-1}$, where N_a and N_b are the number of A and B present and V is the volume of the system, here chosen to be roughly the volume of a bacterial cell (0.7 μm^3). The reverse reaction would have a probability per unit time of $P = N_c \cdot K_b\ s^{-1}$. Kinetic rates for CP–CP interactions are estimated from $K_b/K_f = K_d = \exp(-G/kT)$, where G is proportional to the number of CP–CP contacts made and K_f is fixed at 10^6 per second as in ref. 1. For reactions involving RNA–CP contacts, we estimate rates from known RNA–CP binding assays. In these experiments, the highest affinity RNA–CP interaction was measured to have a K_d of 1.5 nm. Using $K_d = \exp(-G/kT)$, we estimate the upper limit of binding free energy for an RNA–CP interaction to be 12.0 kcal/M. In addition, the lifetime of this RNA–CP interaction was also measured and has been estimated to be on the order of 45–60 s, corresponding to a kinetic off rate of $k_b = 0.1\ s^{-1}$. Using $K_d = k_b/k_f$, we compute the forward kinetic rate of this reaction to be $k_f = 1.1 \times 10^7$ per second per mole. This k_f is used for all RNA–CP interactions and off rates are determined via the RNAs binding free energy for CP, which is allowed to vary from −4.0 kcal/M to −12.0 kcal/M (or from K_d = 1.0 mM to K_d = 1.5 nM). Once the reaction probabilities are computed, we choose a reaction to fire according to their relative probabilities. For this, a random number r is chosen between 0 and 1 and the P_i for each reaction are summed up until the partial sum $P_{partial}$ = Sum_i = 1, n P_i exceeds the sum P_{tot} of all P weighted by r, i.e., until $P_{partial} > P_{tot} \cdot r$. The reaction n at which this occurs is then fired. Thus, reactions with higher probabilities relative to the other reactions will have a higher chance of being fired. After selection of the reaction to fire, the time increment is computed according to $\tau = Ln(1/r)/P_{tot}$, where r is again a random number between 0 and 1.

1. Dykeman EC, Stockley PG, Twarock R (2013) Building a viral capsid in the presence of genomic RNA. *Phys Rev E Stat Nonlin Soft Matter Phys* 87(2):022717.

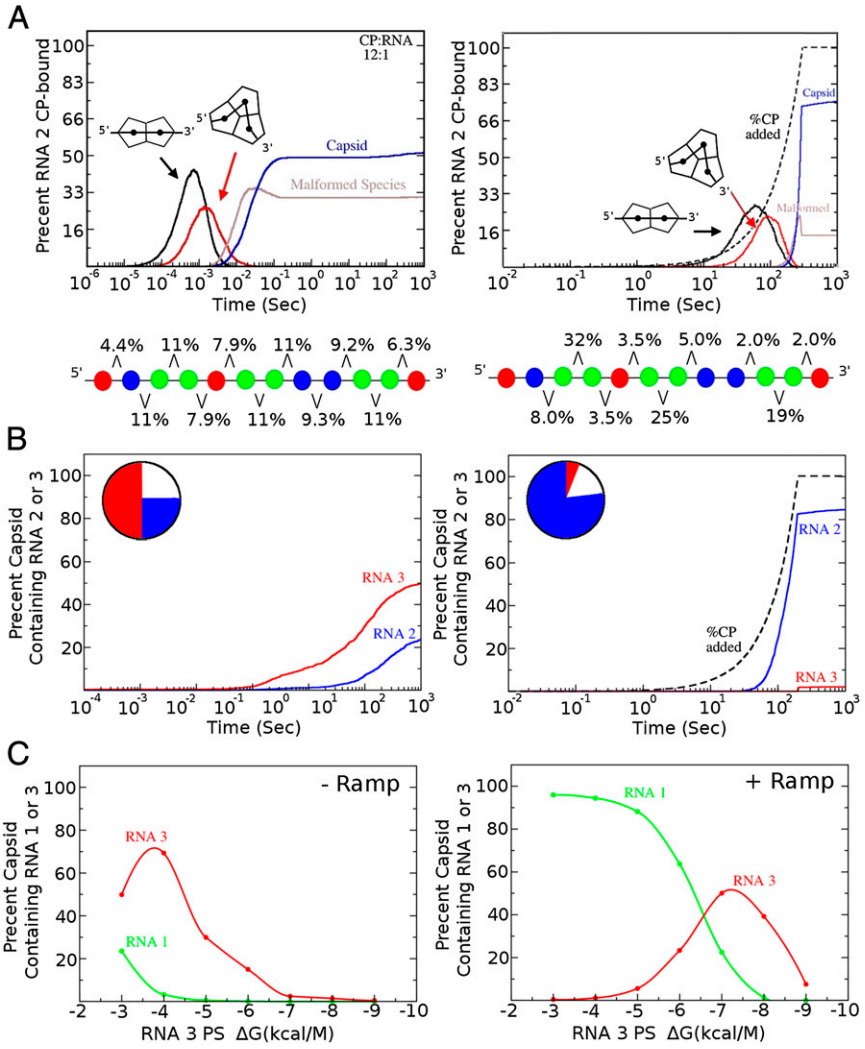

Fig. S1. Different scenarios for the in silico competition experiment. (*A*) Assembly kinetics of the worst assembling RNA, RNA2 (Fig. 2*A*), details as in Fig. 2*B*. (*B*) Competition between RNA2 and RNA3 at a concentration ratio of 1:300 shows a similar outcome to the corresponding experiment with RNA1 in Fig. 4, albeit with a reduced viral load of ~80% in the presence of the ramp. (*C*) Amounts of viral RNA (RNA1) and cellular RNA (RNA3) packaged, as a percent of the total possible particles, as the packaging signal affinities of the cellular RNAs vary. The ramp is essential for avoiding misincorporation. In the presence of the ramp, and RNA1 is the dominant species packaged for a wide range of affinities, showing that this effect is robust.

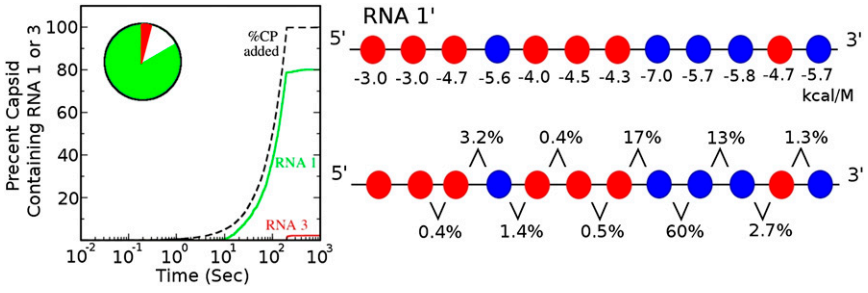

Fig. S2. Consequences of disrupting assembly initiation sites via mutagenesis. The importance of the highest-affinity PSs in RNA1 was assessed by reducing their affinities to −3 kcal/M, producing RNA1′. The absence of dimer and trimer intermediates (Fig. 2*B*), in addition to the assembly kinetics showing that capsid assembly occurs at a similar rate to CP addition, indicates that capsid assembly occurs immediately upon the first binding of CP to an RNA. This suggests that the strong PSs are important for a controlled nucleation event in which the strong PSs sequester CP in the early phases of CP production, preventing individual RNAs from completing until most RNAs have been nucleated. The effect of the changes in the affinities of the first two packaging signals on the nucleation of assembly is shown on the sequence underneath RNA1′, which represents the packaging signals of RNA1′ together with an indication of the relative frequencies at which nucleation occurs at different neighboring PSs (Fig. 2*B*). Nucleation never occurs at the first pair (at the 5′ end, as was the case for RNA1), and the dominant locus of nucleation (60%) has now moved to positions 8 and 9, but these sites have only moderate affinity for CP. The deleterious consequences of these changes for assembly initiation suggest that targeting such PS–CP interactions could be therapeutically beneficial.

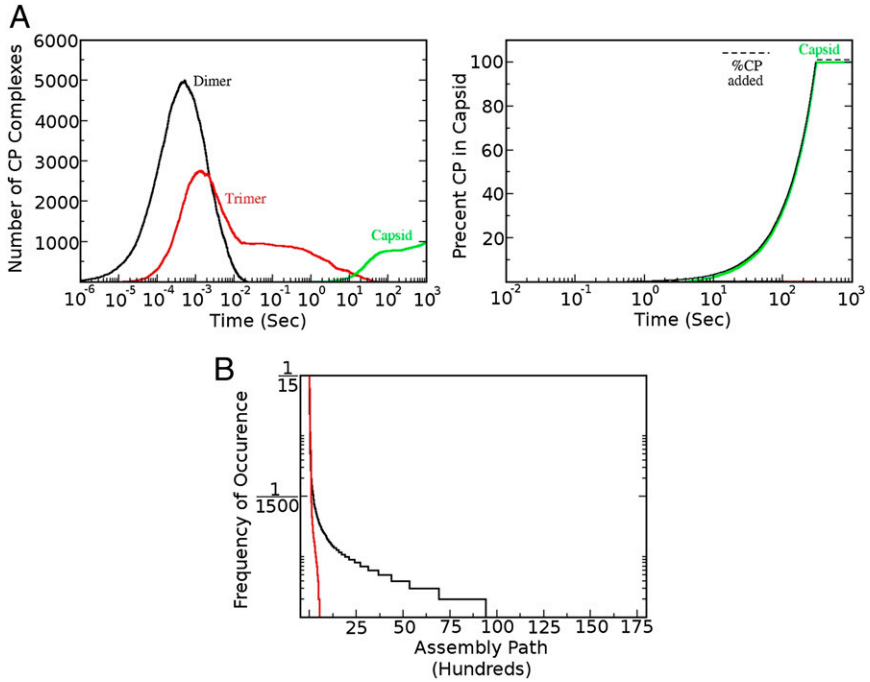

Fig. S3. Container assembly in the absence of genomic RNA. (*A*) Assembly in the absence of a protein ramp (*Left*) shows kinetically trapped species and a low assembly yield; by contrast, 100% assembly is achieved in the presence of the ramp. Assembly has been simulated for a CP–CP binding energy of −4.75 kcal/M. Note that the binding energy has been increased, compared with the −3.75 kcal/M used in the simulation shown in Fig. 2, to illustrate the trapping effect for the protein-only scenario. This value is within the range of CP–CP binding energies that have been previously reported in other work (1–5). (*B*) As before, the protein ramp biases assembly toward a subset of the possible assembly paths. Note that in the protein-only case, the number of possible assembly paths is larger as is reflected by the higher numbers on the *x*-axis in comparison with Fig. 3.

1. Li C, Wang JC, Taylor MW, Zlotnick A (2012) In vitro assembly of an empty picornavirus capsid follows a dodecahedral path. *J Virol* 86(23):13062–13069.
2. Zlotnick A (1994) To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J Mol Biol* 241(1):59–67.
3. Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ (2000) Mechanism of capsid assembly for an icosahedral plant virus. *Virology* 277(2):450–456.
4. Endres D, Zlotnick A (2002) Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* 83(2):1217–1230.
5. Lago H, Parrott AM, Moss T, Stonehouse NJ, Stockley PG (2001) Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: Identification of a protein conformer unable to bind RNA. *J Mol Biol* 305(5):1131–1144.