Published in final edited form as:

J Mol Biol. 2015 July 31; 427(15): 2451–2467. doi:10.1016/j.jmb.2015.05.008.

The role of packaging sites in efficient and specific virus assembly

Jason D. Perlmutter and Michael F. Hagan*

Martin Fisher School of Physics, Brandeis University, Waltham, MA, USA

Abstract

During the lifecycle of many single-stranded RNA viruses, including many human pathogens, a protein shell called the capsid spontaneously assembles around the viral genome. Understanding the mechanisms by which capsid proteins selectively assemble around the viral RNA amidst diverse host RNAs is a key question in virology. In one proposed mechanism, short sequences (packaging sites) within the genomic RNA promote rapid and efficient assembly through specific interactions with the capsid proteins. In this work we develop a coarse-grained particle-based computational model for capsid proteins and RNA which represents protein-RNA interactions arising both from nonspecific electrostatics and specific packaging site interactions. Using Brownian dynamics simulations, we explore how the efficiency and specificity of assembly depend on solution conditions (which control protein-protein and nonspecific protein-RNA interactions) as well as the strength and number of packaging sites. We identify distinct regions in parameter space in which packaging sites lead to highly specific assembly via different mechanisms, and others in which packaging sites lead to kinetic traps. We relate these computational predictions to *in vitro* assays for specificity in which cognate viral RNAs compete against non-cognate RNAs for assembly by capsid proteins.

Graphical abstract

*hagan@brandeis.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

I. INTRODUCTION

In many single-stranded RNA virus families, the spontaneous assembly of a protein container (capsid) around the viral RNA is an essential step in the viral life cycle [1]. Formation of an infectious virion requires that the assembling proteins select the viral RNA out of the milieu of cellular RNA, and most viruses do so with high specificity (e.g. 99% [2]) *in vivo*. Understanding the mechanisms which enable such specific co-assembly could guide the design of delivery vectors that assemble around specific drugs or genes, and could identify targets for antiviral agents that interfere with genome packaging. In this work, we use dynamical computer simulations to investigate the ability of sequence-specific RNA-protein interactions (packaging sites) to drive selective packaging of the viral genome, and how specificity depends on the underlying sequence-independent interactions.

A key driving force for RNA-capsid protein co-assembly is provided by electrostatic interactions between RNA phosphate groups and basic amino acids, often located in flexible tails known as arginine rich motifs (ARMs) [3]. These nonspecific interactions are sufficient for assembly, as shown by the ability of capsid proteins to assemble *in vitro* around heterologous RNA, synthetic polyelectrolytes, and other negatively charged substrates [4–17]. *In vitro* assembly assays [18] and computational modeling [19, 20] indicate that the charge and structure arising from base pairing of viral RNAs is optimal for assembly by their capsid proteins. However, these physical characteristics alone cannot explain the remarkably specific packaging of the viral genome achieved by many RNA viruses *in vivo*. Several factors have been proposed to explain specific packaging *in vivo*, including subcellular localization of viral components [21], coordinated translation and assembly [22–24], and NA-sequence-specific interactions between capsid proteins and sites within the genome called packaging sites (PSs). PSs have been identified for a number of unrelated viruses infecting plant, animal, or bacterial hosts, suggesting this mechanism has widespread relevance [25–33].

The specificity conferred by PSs has been explored through *in vitro* experiments, either by comparing assembly yields of capsid proteins around cognate (i.e. PS containing) and non-cognate RNAs in separate experiments, or by competition assays, in which two RNA species

compete for packaging under limiting protein concentrations. Measured selectivities have varied widely, ranging from high selectivity for the cognate [30, 31, 34], no selectivity [35], or selectivity for a non-cognate RNA [18]. Two recent experiments observed that assembly around cognate RNAs proceeded via different, faster assembly pathways than around non-cognate RNAs [36, 37]. The authors suggest that their experiments are more selective for cognate RNAs because they use a lower protein concentration than previous experiments (50 nM – 1 μ M vs. 10 μ M).

Using chemical kinetics simulations (Gillespie algorithm [38–41]), Dykeman et al. [22, 42] predicted that assembly under dynamic subunit concentrations, i.e. the concentration increase ('ramp') that occurs during an infection cycle in *E. coli* [43], could lead to 100% specificity for RNAs with PSs (represented by nonuniform protein binding affinities) even under a large excess of non-cognate RNAs (represented by uniform binding affinities). In contrast, constant subunit concentrations led to weak differences in yield (~ 5%) and a significant portion of malformed capsids. However, these simulation results do not entirely address the recent *in vitro* experiments [36, 37] in which PSs led to high yield assembly while non-cognate assembly was unsuccessful using constant subunit concentrations.

We recently developed a particle-based computational model for RNA and capsid proteins [19, 44], allowing us to simulate capsid assembly without preassuming the set of allowed assembly intermediates (see section II for further comparison with Gillespie simulations). Although the model is coarse-grained, model predictions for RNA lengths that optimize capsid thermostability quantitatively agreed with viral genome length for seven viruses [19]. We previously examined how varying the nonspecific electrostatic RNA-protein subunit interactions, solution conditions, and subunit-subunit interactions leads to a range of assembly outcomes and different classes of assembly pathways [44].

Here, we explore how introducing specific PS interactions, in a simple form inspired by a recent structural investigation of STNV [37], alters these assembly pathways and products. By extensively comparing assembly around uniform polyelectrolytes (representing non-cognate RNA) and PS-containing polyelectrolytes (cognate RNA), we identify solution conditions that lead to highly specific packaging of the cognate RNA. Depending on the relative strength of protein-protein and protein-RNA interactions, we find that PSs can drive specific assembly via several mechanisms. Consistent with recent single molecule experiments [36], the simulations indicate that PSs can trigger assembly via pathways with more compact intermediates as compared to non-cognate RNAs. However, we also find solution conditions under which PSs are unable to drive specific packaging or even lead to kinetic traps. We then investigate how assembly yields and specificity depend on the number and strength of PSs. In general, we find that a combination of one high affinity PS and multiple weak PSs leads to the highest assembly yields, consistent with the identification of multiple weak PSs in viral genomes [26] and with previous observations that productive self-assembly reactions require reversible interactions [45, 46]. We conclude by discussing potential experimental predictions suggested by these simulations.

II. MODEL

To study the effect of PSs on assembly, we have extended a recently developed model [19, 44] for assembly around linear polyelectrolytes and non-cognate RNAs to include a representation of PSs. The model is motivated by recent experiments in which purified simian virus 40 (SV40) capsid proteins assemble *in vitro* around ssRNA molecules to form virus-like particles composed of 12 homopentamer subunits [47, 48]. The model capsid is therefore a dodecahedron comprising 12 pentagonal subunits, each of which represents a homopentamer of the capsid protein. It is assumed that homopentamers are stable and form rapidly in solution, as is the case for SV40. Although the structure of the model capsid is motivated by these experiments [47, 48], in this article we use the model to study general relationships between PSs and assembly which could apply to many viral species.

Capsid protein subunit-subunit interactions

Following Refs [19, 44, 49], model subunits are attracted to each other via attractive pseudoatoms, ‘attractors’ (type ‘A’) at the vertices, which interact via a Morse potential (see Fig. 1 and section SIA). The subunit-subunit interaction strength is controlled by the model parameter ϵ_{ss} ; the free energy of subunit dimerization is $g_{ss}/k_B T = 5.0 - 1.5\epsilon_{ss}$. This does not include the effects due to repulsions between ARMs (defined below), which we estimate to reduce g_{ss} by $\sim 0.5k_B T$ at 100mM; see section SIB. This attraction represents the interaction between capsid protein subunits that arises from hydrophobic, van der Waals, and electrostatic interactions [46]), whose strength can be experimentally tuned by pH and salt concentration [1, 46, 50, 51].

Pairs of subunits are driven toward a preferred subunit-subunit angle consistent with a dodecahedron (116 degrees) by repulsive ‘Top’ pseudoatoms (type ‘T’), which interact via the repulsive term of the Lennard-Jones (LJ) potential. The ‘Bottom’ pseudoatoms (type ‘B’) have a repulsive LJ interaction with ‘T’ pseudoatoms, intended to prevent ‘upside-down’ assembly. The ‘T’, ‘B’, and ‘A’ pseudoatoms form a rigid body [49, 52, 53]. See Refs. [53–70] for related models.

Sequence-independent electrostatic interactions

Capsid assembly around nucleic acids and other poly-electrolytes is driven by electrostatic interactions between negative charges on the encapsulated polyelectrolyte and positive charges on capsid protein-RNA binding domains [3, 46]. To account for these interactions, we extend the model as follows. First, we add positively charged bead-spring polymers affixed to the inner surface of the subunit, to represent the highly charged, flexible terminal tails known as arginine rich motifs (ARMs) that are typical of positive-sense ssRNA protein-RNA binding domains [3]. There are five ARMs per pentameric subunit. For each ARM, the first segment is anchored at a fixed position on the subunit, midway between the subunit center and a vertex. Except where stated otherwise, each ARM contains five segments of charge $+e$. To better represent the capsid shell, we include a layer of ‘Excluder’ pseudoatoms, which have a repulsive LJ interaction with the RNA and the ARMs. The ‘Excluders’ and first ARM segment are part of the subunit rigid body. ARM beads interact through repulsive LJ interactions and, if charged, electrostatic interactions.

To represent an RNA molecule, we consider a linear bead-spring polyelectrolyte, with a charge of $-e$ per bead and a persistence length comparable to that of ssRNA in the absence of base pairing. To focus on the effect of PSs, we do not consider RNA base pairing in this work; the effect of base pairing on assembly was considered in Ref. [19]. We also previously determined how assembly depends on polyelectrolyte length [19]. In the present work, for each simulated salt concentration and capsid structure, we use the polyelectrolyte length that optimizes assembly around non-cognate RNAs. The values of optimal lengths as a function of salt concentration are shown in Fig. S8. Except in extreme cases, the introduction of PSs only slightly changes the optimal lengths.

Electrostatics are modeled using Debye-Hückel (DH) interactions, where the Debye screening length (λ_D) is given by $\lambda_D \approx 0.3/C_{\text{salt}}^{1/2}$ with λ_D in nm and C_{salt} the concentration of monovalent salt in molar units. Perlmutter et al. [19] showed that DH interactions compare well to simulations with explicit counterions for the parameter values under consideration; comparisons between simulations with DH interactions and those with explicit counterions are presented in Refs. [19, 44] and Fig. S8.

Packaging sites (PSs)

Structures of PSs obtained for a number of viruses through x-ray crystal structures and/or bioinformatics correspond to short stem loops [26]. For example, multiple short stem-loops with a single-stranded loop motif of A.X.X.A-, where X corresponds to any nucleotide, were identified as PSs in the STNV genome [32]. An x-ray structure of STNV VLPs containing stem loop fragments revealed that they bind to well-defined sites on the protein ARMs with the effect of bringing multiple subunits into proximity and favoring alignments conducive to subunit-subunit interaction, thus enhancing subunit-subunit interactions as well as subunit-RNA interactions [37].

To account for these effects, we have extended the model to include a generic representation of PS interactions by adding new pseudoatoms (denoted as packaging site receptors, PSRs) to the model protein subunits (Fig. 1B). The PSRs experience short-range interactions with particular RNA segments that correspond to PSs. For simplicity, the PS-PSR interaction uses the same short-range Morse potential as the attractor-attractor interaction (Eq. S4, section SIA). The strength of the PS-PSR attractive interaction is parameterized by the interaction well-depth ϵ_{PS} . There are five PSRs per pentameric subunit. Except where noted otherwise, each PSR is located approximately midway between an ARM anchor segment and a subunit vertex. This location allows for a PS to simultaneously bind to 3 PSRs when three subunits form an optimal geometry. Thus, the PSs not only promote subunit-RNA binding, but also generate RNA-mediated subunit-subunit interactions, as inferred from structural data [37].

In our simulations, we explore how assembly depends on **(A)** the number of PSs (N_{PS}), **(B)** the PS binding affinity (ϵ_{PS}), and **(C)** distribution of PS binding affinities along the polyelectrolyte.

- A.** There are 60 PSRs in a complete model capsid, located at 20 threefold axes. Thus, an RNA with 20 PSs can interact with every PSR, and we will refer to $N_{\text{PS}} = 20$ as

the ‘stoichiometric’ number of PSs. Since recent experiments have identified a range of values for N_{PS} in viral genomes (*e.g.* [32, 37, 71–74]), we perform simulations with $N_{\text{PS}} \in [0, 40]$.

- B. To limit the number of model parameters, we consider two classes of model PSs: high affinity sites, with PS-PSR interaction well depth $\varepsilon_{\text{PS}} = 20k_{\text{B}}T$ (see Eq S11) and low affinity sites, with $\varepsilon_{\text{PS}} = 5k_{\text{B}}T$. We calculated a μM dissociation constant for the low affinity PS (see Section SIB), consistent with experimental observations [75].
- C. To describe how assembly depends on the distribution of affinities, we consider three forms of PS distributions along the model RNA: (i) N_{PS} high affinity PSs, (ii) N_{PS} low affinity PSs, and (iii) 1 high affinity PS along with N_{PS} low affinity PSs. In each case the PSs are placed at a uniform interval along the RNA. Unless otherwise noted, the strong PS in distribution *iii* is placed at the center of the RNA as found for MS2 [31].

Simulations, units, parameters, and statistics

Simulations were performed with the Brownian Dynamics algorithm of HOOMD, which uses the Langevin equation to evolve positions and rigid body orientations in time [76–78]. Simulations were run using a set of fundamental units. The fundamental energy unit is selected to be $E_{\text{u}} \equiv 1k_{\text{B}}T$. The unit of length D_{u} is set to the circumradius of a pentagonal subunit, which is taken to be $1D_{\text{u}} \equiv 5 \text{ nm}$ so that the dodecahedron inradius of $1.46D_{\text{u}} = 7.3 \text{ nm}$ gives an interior volume consistent with that of the smallest $T=1$ capsids. To calculate the thermodynamic optimal encapsidation length, we placed a very long polymer in or near a preassembled capsid, with one of the capsid subunits made permeable to the polymer and performed unbiased Brownian dynamics. Once the amount of packaged polymer reached equilibrium, the thermodynamic optimum length was measured. We previously [19] found that this strategy closely matched that produced using the Widom insertion method [79] as applied to growing polymer chains [67, 80].

Assembly simulations were run at least 10 times for each set of parameters, each of which were concluded at either completion, persistent malformation, or 2×10^8 time steps. This observation time was chosen based on the time after which assembly yields and outcomes change only logarithmically with time for most parameter values. For all assembly simulations there were 60 subunits, *e.g.* enough to form five complete capsids, with box size $= 200 \times 200 \times 200 \text{ nm}$, resulting in a concentration of $12\mu\text{M}$. Due to the limited number of simulations we are able to run at each parameter combination, we use bootstrapping to estimate statistical errors.

A summary of all model parameters and their connection to biological or *in vitro* assembly experiments is presented as Table I. Finite computational resources place limitations on our model; for example, we have restricted our simulations to a small, $T=1$ capsid. We also note that, while our model was originally intended as a general model of viral self-assembly that can be adapted to represent a variety of specific viruses [19], the specific motivation for the

PS interaction added here comes from STNV. Thus, wherever possible, we use references that study STNV assembly in Table I.

By comparing the diffusion constant of subunits within our simulations with the expected diffusion constant for a particle with the size of an SV40 pentamer in solution, we estimate total simulated timescales as approximately 10ms. *In vitro* experiments with SV40 observe capsids assembling within tens of ms [47], while experiments with STNV and MS2 found assembly to occur on the timescale of minutes [36]. Due to computational limitations we have focused on parameters that lead to assembly on the shortest applicable timescales. However, we have previously used Markov state modeling to show that, under other parameter values, assembly rates are two orders of magnitude slower (i.e. seconds). [19]. Also note that such estimates of timescales should only be taken as qualitative. By reducing the degrees of freedom, coarse-grained models such as employed here typically generate energy landscapes which are less rugged than those of atomistic models, which can significantly speed up dynamics [81]. Moreover, we are not accounting for processes such as capsid protein conformational transitions and RNA folding. While the rate at which RNA can locally fold to form PSs is thought to be significantly faster than the assembly timescale [82], how RNA secondary and tertiary structures respond to the presence of capsid subunits during assembly remains a subject of research [3].

VMD was used to visualize the model conformations [83] and the hydrodynamic radius R_H was calculated using HYDROPRO [84] as discussed previously [44].

Comparison to existing models

Dykeman et al. [22, 42] extended the kinetic rate equation approach of Becker and Doring [85] and Zlotnick [86] to include a representation of RNA and PSs, simulated using the BKL or Gillespie algorithm [38–41]. In Gillespie algorithm simulations, the state space (the set of allowed partial capsid geometries and RNA configurations) and the transition rates (e.g. association rates among RNA-bound subunits) must be assumed *a priori* [38, 39, 54]. Since we seek to determine the ensemble of possible assembly pathways and products in this work, we employ Brownian dynamics simulations in which spatial motions of particles are explicitly resolved. We note that our simulations require *a priori* specification of interparticle interactions, but collective processes such as cooperative RNA-protein motions [67] emerge from the dynamics without additional assumptions. The different limitations of Gillespie algorithm and Brownian dynamics simulations are discussed in Ref. [46].

Dykeman et al. present models which assume that each of the PSRs is bound by RNA once and that subunits adding to partial capsids are bound to adjacent RNA segments, so that assembly must follow a Hamiltonian path [22, 42]. In the present model the spatial positions and dynamics of subunits are explicitly tracked and thus there are no assumptions made about the state space or assembly pathways. Consequently, there are no explicit restrictions on the sequence of RNA binding sites, although steric hindrances disfavor binding of multiple PSs at the same threefold axis and RNA conformational statistics favor returning to nearby binding sites.

III. RESULTS

In this section we describe how assembly depends on the specific (PS-PSR) and nonspecific (electrostatic) interactions. We refer to polyelectrolytes equipped with only nonspecific interactions as **non-cognate** RNAs, and polyelectrolytes which contain one or more PSs as **cognate** RNAs. Note that since we neglect base pairing in this work, the non-cognate RNA is a linear polyelectrolyte. However, since we also neglect base pairing in the cognate RNA, our results are applicable to a comparison between cognate and non-cognate RNAs, except to the extent that the tertiary structure of cognate RNAs is more favorable for assembly than that of non-cognate RNAs [19, 87].

A. The yield and specificity conferred by PSs depends on subunit-subunit and nonspecific electrostatic interactions

Yield without PSs—Assembly around non-cognate RNA (uniform polyelectrolytes) depends on the strength of subunit-subunit interactions (controlled by ϵ_{ss} in our model, eq. S2 in section SIA, and salt concentration or *pH in vitro*) and sequence-independent electrostatic interactions (controlled by the salt concentration C_{salt}). The dependence of assembly outcomes on these parameters for a non-cognate RNA is summarized in Fig. 2. High yields of well-formed VLPs are observed for $C_{salt} \in [50 - 400]$ mM and moderate subunit-subunit interaction strengths, $\epsilon_{ss} \in [4 - 6]k_B T$. Outside of optimal parameter values, yields are suppressed by several failure modes: strong electrostatics lead to disordered aggregates, strong subunit-subunit interactions lead to malformed capsids, and overly weak interactions lead to unnucleated complexes [44]. The yield $f_{nc}(t)$ is defined as the fraction of simulations which, at time t , result in formation of a complete capsid (defined as 12 subunits each strongly interacting with five neighbors) completely encapsulating the non-cognate RNA. The yield $f_{nc}(t_{end})$ at the simulation endpoint t_{end} is shown in Fig. 2D.

Yield with PSs—Based on observations of multiple low affinity PSs [26] and simulations at varying numbers and strengths of PSs (section III B below), we performed simulations at varying ϵ_{ss} and C_{salt} for an RNA with 1 high affinity PS and $N_{PS}=25$ low affinity PSs (see section SIA). With the addition of PSs, the range of parameters leading to high assembly yields (f_c) broadens considerably (Fig. 2B,E), allowing assembly at much lower values of ϵ_{ss} across a wide range of C_{salt} and increasing the upper range of ϵ_{ss} leading to assembly at low C_{salt} .

Notably, assembly around the non-cognate RNA fails in distinct ways in these different parameter regions (Fig. 2C), indicating that PSs can avoid multiple forms of thermodynamic or kinetic traps. At low ϵ_{ss} , the predominant effect of PSs is to enhance nucleation and growth rates by increasing effective interaction strengths. Increased subunit-RNA interactions are most relevant at high salt, while increased subunit-subunit interactions are most relevant at low salt (discussed in section III B). At high ϵ_{ss} and low or moderate salt (*e.g.* $C_{salt}=50$ mM and $\epsilon_{ss}=7k_B T$) assembly around non-cognate RNA frequently leads to the nucleation of multiple partial capsids on the same RNA; typically these intermediates have incompatible geometries and either fail to combine or form malformed capsids. In the cognate RNA simulations, assembly rapidly nucleates around the HA PS; the LA PSs then

enhance growth rates such that assembly is completed before additional partial capsids can nucleate elsewhere on the RNA. The frequency of multiple nucleation events as well as the structural heterogeneity of assembly intermediates are shown in SI Fig. S4.

Specificity—An estimate of the specificity conferred by PSs can be obtained by comparing the assembly dynamics in the presence and absence of PSs. We calculated the probability that, for a given ε_{ss} and C_{salt} , assembly of a well-formed capsid occurs around the cognate RNA before the non-cognate RNA, normalized by the probability of complete assembly around either substrate:

$$P_1 = \frac{\int_0^{t_{end}} dt P_c(t) [1 - f_{nc}(t)]^{r_{ex}}}{f_c(t_{end}) + (1 - [1 - f_{nc}(t_{end})]^{r_{ex}}) - f_c(t_{end}) (1 - [1 - f_{nc}(t_{end})]^{r_{ex}})} \quad (1)$$

where $f_{nc}(t)$ and $f_c(t)$ are the time-dependent yields around non-cognate and cognate RNAs (measured from the simulations whose final yields are shown in Figs. 2D,E), and

$P_c(t) = \frac{df_c(t)}{dt}$ is the assembly time probability distribution function for cognate RNAs. The parameter $r_{ex} = c_{nc}/c_c$ is the ratio of non-cognate to cognate RNAs. The result of applying Eq. 1 to the data in Fig. 2D,E for $r_{ex}=1$ is shown in Fig. 3A.

In a fairly wide range of parameter space, the assembly is 100% specific for assembly around the cognate RNA. However, at parameters which are optimal for assembly around non-cognate RNA (i.e. where assembly without PSs leads to high-yield, $\varepsilon_{ss} \in [4 - 6]k_B T$, $C_{salt} \in [100 - 300]mM$), there is essentially no selectivity. This result highlights the importance of the solution conditions when assessing the role of PSs *in vitro*, and may suggest an explanation for the varying levels of specificity for cognate RNAs observed in *in vitro* experiments (see the Introduction).

We next compare this competition estimate approach to explicit competition simulations which contain a cognate RNA, r_{ex} non-cognate RNAs, and 60 pentamer subunits. In these simulations, we define specificity as the fraction of simulations in which the first assembled capsid forms around a cognate RNA. Figure 3B presents results for $r_{ex} = 1$ at several subunit-subunit interaction strengths and salt concentrations. For several parameter sets in that figure, assembly is not productive without PSs, and so as expected selectivity is 100%. At the other three parameter sets, which result in incomplete selectivity, the predicted and measured values agree to within error.

While *in vitro* assays have typically focused on competition between equal concentrations of cognate and non-cognate RNAs, assembly *in vivo* can occur under a large excess of cellular RNAs [43]. The estimated specificity for $r_{ex}=10$ is shown in Fig. 3C. We see that excess non-cognate RNA strongly suppresses encapsulation of the genome under parameter values where assembly around the non-cognate RNA proceeds at a similar rate. However, selective genome encapsulation still occurs under parameter values where assembly around the genome is strongly favored. Extrapolating to large excess (e.g. $r_{ex}=300$ such as can occur within *E. coli*), selectivity for the cognate would be restricted to the lower left regions of Figures 3A and 3C, where assembly around the genome is orders of magnitude more rapid than around non-cognate RNA.

We have also considered competition under subunit limiting conditions — 1 cognate RNA, r_{ex} non-cognate RNAs, and 18 protein subunits — so that at most one complete capsid can assemble, containing either a cognate or a non-cognate RNA. As expected based on the independent assembly simulations, for parameters where assembly is not productive without PSs ($C_{\text{salt}} = 500\text{mM}$, $\varepsilon_{\text{ss}} = 6k_{\text{B}}T$), we observe that assembly is 100% specific for the cognate RNA in all simulations for $r_{\text{ex}} \in [1 - 50]$. Interestingly, assembly is also 100% specific for the cognate RNA at a parameter set for which assembly is productive without PS ($C_{\text{salt}} = 500\text{mM}$, $\varepsilon_{\text{ss}} = 7k_{\text{B}}T$, $r_{\text{ex}} = 1$). For these parameters, nucleation occurs first around the PS, which reduces the number of free subunits and impedes nucleation around the non-cognate. We are exploring whether this effect remains in larger systems.

We note that the relationship between P_1 and specificity in direct competition assays could break down at low salt, where subunits initially undergo nonspecific absorption onto cognate and non-cognate RNAs. Assembly under limiting subunit concentrations in these conditions requires exchange of subunits between RNAs, which occurs slowly relative to our simulation timescales.

A final observation is that the degree of specificity observed in our simulations does not correspond to the relative Boltzmann weights for cognate and non-cognate encapsidation. We estimate that the PSs within the Combo sequence contribute $\sim -400k_{\text{B}}T$ to the free energy of encapsidation (using Fig S1), which strongly favors encapsidation of the cognate at equilibrium for $r_{\text{ex}}=1$ for all values of C_{salt} and ε_{ss} . Yet, where assembly around non-cognate RNAs proceeds rapidly and with high-fidelity, we observe $P_1 \sim 0.5$ despite the strong thermodynamic favorability of cognate encapsidation. This result exemplifies the competition between thermodynamics and kinetics that takes place in any finite-time assembly reaction [45, 46].

B. The effect of PSs depends on their number and strength

We now discuss the dependence of assembly and specificity on the number and affinities of PSs. We focus on two interaction parameter sets: $C_{\text{salt}}=100\text{mM}$, $\varepsilon_{\text{ss}}=2k_{\text{B}}T$ and $C_{\text{salt}}=500\text{mM}$, $\varepsilon_{\text{ss}}=6k_{\text{B}}T$. These parameters lead to 100% specificity for the cognate sequence considered in Fig. 2 (1 high affinity PS and $N_{\text{PS}}=25$ low affinity PSs), but represent very different strengths of nonspecific electrostatic interactions, and correspondingly different assembly pathways around non-cognate RNAs. For each of these interaction parameter sets, we simulated three distributions of PS affinities along the model RNA: (i) N_{PS} high affinity (HA) PSs, (ii) N_{PS} low affinity (LA) PSs, and (iii) a ‘Combo’ distribution with 1 HA PS and N_{PS} LA PSs. Recall that the Combo sequence with $N_{\text{PS}}=25$ is considered in Figs. 2 and 3.

As shown in Fig. 4, assembly yields for both interaction parameter sets are most robust under the Combo PS distribution. High yields are obtained for intermediate values of N_{PS} , although the yield is optimal for sub-stoichiometric $N_{\text{PS}} < 20$ at moderate salt and super-stoichiometric $N_{\text{PS}} > 20$ at high salt (recall that there are $N_{\text{PS}} = 20$ PS binding sites in a complete capsid).

However, the effect of PSs on assembly mechanisms, and hence the dependence on PS distribution, is markedly different for the two parameter sets. At $C_{\text{salt}}=100\text{mM}$, the subunits

rapidly adsorb onto the RNA. Without PSs, the weak subunit-subunit interactions ($\epsilon_{ss}=2k_B T$) are insufficient to drive subsequent assembly resulting in a disordered aggregate (Fig. 2A). Nonetheless, even a sub-stoichiometric number of PSs is sufficient to promote complete assembly (Fig. 4A). High yields are observed for 6–8 HA PSs, 10 LA PSs, and for $N_{PS} \in [10, 20]$ for the Combo case. For larger than optimal N_{PS} , multiple partial capsids nucleate on the same RNA, leading to long-lived malformed assemblies that suppress yields. Snapshots illustrating typical assembly outcomes at low, stoichiometric, and excess N_{PS} are shown below the plots for each salt concentration in Fig. 4. For the Combo PS sequences, after several subunits have adsorbed onto the RNA, the strong PS initiates assembly, with further growth mediated by the weak PS. When there are multiple HA PSs, it is more likely for multiple small clusters to form, which may then merge into a single capsid, with the final additions driven by electrostatic interactions.

We note that the effect of PSs under low salt and low ϵ_{ss} derives not from their ability to drive subunit-RNA interactions, which are already strong due to nonspecific electrostatics, but rather because the locations of packaging site receptors (PSRs) at the capsid three-fold axes promotes subunit-subunit interactions (Fig. 1B). In support of this conclusion, simulations in which PSRs were moved to the center of model subunits led to poor assembly (♦ symbol in Fig. 4A).

At high salt concentration ($C_{salt}=500$ mM), different PS sequences promote assembly (Fig. 4B). Without PSs under these conditions, few subunits absorb on the RNA and nucleation does not occur on the timescales being simulated (using Markov State modeling we determined that assembly eventually occurs around the non-cognate RNA on a timescale which is two orders of magnitude longer [44]). With sub-stoichiometric N_{PS} , a cluster of subunits assembles in the vicinity of PSs, but subsequent growth into a capsid is slow on simulated timescales. In contrast, moderate to high yields are observed for super-stoichiometric PSs ($N_{PS} \in [30 - 40]$). For the Combo sequence, the HA PS promotes rapid nucleation of a trimer after which the LA PSs facilitate adsorption and binding to the cluster by additional subunits. In contrast to the low salt conditions, super-stoichiometric HA PSs also lead to moderate yields of well-formed capsids; because of the weak nonspecific electrostatics malformed capsids are less prevalent.

With high salt and relatively strong subunit-subunit interactions ($\epsilon_{ss}=6k_B T$) the ability of PSs to drive subunit-RNA interactions should be most relevant to promoting assembly. The effective subunit-subunit interactions promoted by $N_{PS}=25$ PSs are stronger than optimal, as can be seen by the fact that reducing ϵ_{ss} increases yields (Fig. 2B). Consistent with this reasoning, eliminating the contribution of PSs to subunit-subunit interactions by moving the PSRs to subunit centers increased the yield (♦ symbol in Fig. 4B).

While the location of PSRs within the capsid structure can significantly affect assembly, we found that the location of the strong PS along the RNA (in the Combo sequence) did not measurably alter the yield (open pentagon symbols in Figs. 4A,B). This observation approximately agrees with Dykeman et al. [42] who predicted a very weak dependence on HA PS location.

C. PSs can alter assembly pathways

Modeling [44, 66, 88–91] and experiments [47, 92, 93] have shown that assembly pathways around non-cognate RNAs can be classified according to two extremes. Systems in which protein-protein interactions dominate assemble through nucleation-and growth pathways with ordered intermediates, whereas strong protein-RNA interactions (low salt and/or high ARM charges) lead to the ‘en masse’ mechanism in which subunits rapidly adsorb on the RNA in a disordered manner, followed by cooperative rearrangements to form a capsid. Assembly pathways can be classified by the parameter n_{free} , defined as the number of subunits adsorbed to the RNA which are not part of the largest subunit cluster, averaged over system configurations for which the largest partial capsid intermediate has 4–6 subunits [44]. For our model capsid with 12 subunits, $n_{\text{free}} \gtrsim 5$ indicates the en masse mechanism, with smaller values indicating the nucleation-and-growth mechanism. As shown in Fig. 5, assembly pathways around our model non-cognate RNA range the gamut of n_{free} , with low salt and low ϵ_{ss} leading to en masse pathways and high salt and high ϵ_{ss} leading to nucleation and growth pathways.

The addition of PSs has a striking effect on assembly pathways. As shown in Fig. 5, assembly pathways around the Combo PS sequence with $N_{\text{PS}} = 25$ correspond to the nucleation-and-growth mechanism over a broad range of ϵ_{ss} and C_{salt} . Under most conditions the PSs increase the order of assembly intermediates (lowering n_{free} , Fig. 5C) because adsorbed subunits are co-localized and well-positioned for assembly. However, under high ϵ_{ss} the most significant effect of PSs is to increase subunit adsorption on the RNA and thus PSs slightly increase n_{free} . Snapshots from representative trajectories for both of these cases are shown in Fig. 6. Consequently, n_{free} and correspondingly the nature of assembly pathways are less sensitive to conditions (ϵ_{ss} and C_{salt}) than for the non-cognate RNA. This result parallels the observation that PSs reduce the sensitivity of assembly yields to control parameters (Fig. 2).

Relationship between predicted assembly pathways and single molecule fluorescence correlation spectroscopy (smFCS) data—

A means to test the predicted dependence of assembly pathways on solution conditions and PSs is provided by the fact that pathways with different values of n_{free} can be distinguished by the hydrodynamic radii (R_{H}) of their early intermediates [44]. Recent experiments have used smFCS to monitor the timecourses of R_{H} during assembly around cognate and non-cognate RNAs [36, 94, 95]. Under the experimental conditions, assembly around cognate RNAs was rapid and characterized by either constant R_{H} or a collapsed complex followed by gradual increase to the size of an assembled capsid. Assembly around non-cognate RNAs was slower, with R_{H} initially increasing before finally decreasing to the size of the capsid.

To relate the predicted effect of PS on assembly pathways (n_{free}) discussed above to an experimentally observable quantity, we estimated the hydrodynamic radii R_{H} for polymer-subunit intermediates using the program HYDROPRO, which has been shown to accurately predict R_{H} for large protein and protein-NA complexes [84]. Fig. 7 shows calculated R_{H} for assembly around cognate and non-cognate RNAs for $C_{\text{salt}}=100\text{mM}$ and several values of ϵ_{ss} . For weak subunit-subunit interactions ($\epsilon_{\text{ss}} = 2k_{\text{B}}T$, Fig. 7A), assembly without PSs results in

disordered aggregates (Fig. 2), and the R_H monotonically increases over time. With PSs, on the other hand, the R_H initially increases as subunits attach to the RNA, and then rapidly decreases as the capsid assembles. Upon increasing the subunit-subunit interactions strength ($\epsilon_{ss} = 3k_B T$, Fig. 7B), successful assembly occurs with and without PS; however, the increase in R_H is greater and of longer duration in the absence of PS. This difference occurs because the PSs enhance the assembly rate and decrease n_{free} . Finally, under stronger subunit-subunit interactions ($\epsilon_{ss} = 4k_B T$) PSs have little effect on n_{free} (Fig. 5) and correspondingly the time course of R_H (Fig. 7C) is similar for cognate and non-cognate RNAs.

In the simulations discussed thus far we used a relatively short RNA (575 segments at $C_{salt} = 100mM$), since this is the optimal length for our subunits with the small ARM charge (+5) [19, 44]. Therefore, the R_H of the free RNA (prior to encapsidation) is less than that of the assembled capsid and the measured R_H increases substantially upon adsorption of subunits, until assembly of ordered partial capsids reduces R_H . To examine the applicability of our findings to the more typical case in which the free RNA R_H is similar to or larger than the capsid size, we also performed simulations on subunits with ARMs with charge (+10) and optimal RNA length 910 segments (Fig. 7D–F). The behavior of this system is qualitatively similar to that of the +5 ARMs, except that the initial increase in R_H upon subunit absorption is less apparent.

The results at low subunit-subunit interaction strength and +10 ARM (Fig. 7D) reproduce key features of the experimental observation of STNV assembly — a monotonic increase in R_H for unsuccessful non-cognate assembly, and a rapid decrease in R_H during cognate mediated assembly [36]. Importantly, the ability to change R_H time courses does not depend on specific geometric features of the PSs in our simulations, but only requires that PSs promote ordered assembly pathways. Increasing the subunit-subunit interactions can achieve a similar effect as adding PSs in this regard (Fig. 7E). Note that in previous simulations [44] we found that including base pairing did not qualitatively change R_H time-courses. However, we do not observe an increase in R_H after the initial decrease, as observed in some of the experiments [36]. This pattern might reflect a global RNA conformation change triggered by subunit binding, which is not currently incorporated in our model.

D. Restricted RNA Conformational Dynamics Inhibit Assembly

Although our simulations show that PSs can promote efficient assembly under conditions in which nonspecific assembly fails, we also found that non-optimal numbers and affinities of PSs can lead to kinetic traps. In this section we discuss mechanisms that can lead to these kinetic traps. As seen in Fig. 4, RNAs with 20 HA PSs (the stoichiometric amount) fails to produce capsids at high or low salt. In part, this outcome follows the well-known rule that strong interactions hinder self-assembly into highly ordered low free energy configurations by preventing ‘local’ equilibration among partially assembled configurations and thus trapping the system in metastable disordered states [45, 46, 50, 55, 65, 68, 96–98], as found for strong nonspecific electrostatics ($C_{salt} = 10 mM$) or subunit-subunit interactions ($\epsilon_{ss} = 8k_B T$, Fig. 1). However, a more nuanced explanation is required, since at high salt adding more HA PSs ($N_{PS} = [30, 40]$, $\epsilon_{PS} = 20k_B T$) does lead to moderate assembly yields (Fig. 4).

Analysis of the high salt, $N_{PS} = 20$ simulation trajectories revealed ordered partially assembled capsid intermediates which, despite having optimal subunit-subunit interaction geometries, failed to reach completion (Fig. 8).

We hypothesized that the stalled assemblies result from poorly equilibrated RNA conformational dynamics within assembling capsids. To investigate the dependence of the RNA dynamics on PS sequences, we measured the fluctuations of the paths traced by RNAs within capsids (or capsid intermediates) over the course of dynamical trajectories. To simplify the analysis we define an RNA path as the sequence of capsid threefold sites with which RNA segments interact (counting by the index of the RNA segment which passes within 1 nm of the PS binding site). Because the ARM anchoring sites and the PSRs are located near threefold sites, these sites have enhanced non-cognate RNA densities both in the presence and absence of PSs [19], allowing us to define a path for both the cognate and non-cognate RNAs. We note that these paths can be complex; for example, with strong electrostatics ($C_{salt} = 100mM$), ~ 50% of paths re-visit one or more threefold sites (meaning segments which are nonlocal in sequence interact with the same threefold site), compared with only ~ 10% at ($C_{salt} = 500mM$). Note that in our model only one PS can reside in each PS binding site, consistent with the structure in Ford et al. [37]. Re-visitations are stabilized by nonspecific electrostatic interactions between RNA segments and ARMs at an occupied threefold site. Similarly, jumps between non-neighboring vertices are also common, present in ~ 1/3 of paths. Thus, the RNA usually does not trace a Hamiltonian path within the capsid in our simulations.

In Figure 8 we quantify the RNA path persistence time (time required to change conformation) within assembled capsids. In Fig. 8A, we see that non-cognate RNAs and RNAs with the Combo PS sequences (primarily LA PSs) are highly dynamic; a new RNA path is observed at almost every observation time. In contrast, an RNA with 20 HA PSs is far less dynamic, with existing paths persisting for far longer periods of time. To help visualize the differences between these two classes of dynamics, schematics of pathways observed at five different observation times for different RNA sequences are shown in Fig. 8B. While the paths for the 20 HA PS sequence are nearly identical, the paths change significantly on this timescale for the other two PS sequences. We find that assembly stalls when the RNA becomes frozen in a conformation whose geometry hinders recruitment of additional subunits to the assembling capsid. Two examples of such conformations are shown in Fig. 8C. Under parameters which promote RNA dynamics, such conformations are transient.

Notably, as the number of PSs is increased beyond the stoichiometric value $N_{PS} = 20$ for any distribution, the RNA dynamics increases (Fig. S6). This observation explains the increase in yield for the HA PS distribution at high N_{PS} . This facilitation arises because excess PSs can displace bound PSs from a given three-fold site without requiring complete loss of interaction at that site, thus avoiding the activation barrier associated with PS-PSR unbinding (see SI section SIIB and Fig. S6). In contrast, for $N_{PS} = 20$ exchange requires dissolution of a PS-PSR interaction. In effect, RNA with excess PSs can slide on capsid intermediates to sample different conformations and thus escape from unproductive traps. Previous simulations [67, 99] and a theoretical model [100] have suggested the importance

of RNA rearrangements and subunit ‘sliding’ during assembly around non-cognate polyelectrolytes .

IV. DISCUSSION

In this article, we have described simulations of capsid assembly around RNA, represented as a flexible polyelectrolyte with sequence-specific protein-RNA interactions, or packaging sites (PSs). By performing extensive simulations over a range of salt concentrations, simulated protein-protein interaction strengths, and affinity and number of PSs, we have explored how PSs alter the pathways and products of capsid assembly reactions, and the extent to which they induce specificity against polyelectrolytes without sequence-specific interactions (e.g. non-cognate RNAs). We find that PSs can confer arbitrarily high specificity over RNAs with uniform nonspecific interactions, but that the degree of specificity is sensitive to the underlying assembly driving forces, which can be tuned by solution conditions (ionic strength, pH) as well as capsid protein charge (ARM sequence). Specificity is maximal under conditions where the nonspecific interactions alone are slightly too weak to promote effective assembly.

Assembly and specificity are also sensitive to the affinity and number of PSs, with the optimal distribution of PSs depending on the solution conditions. Our simulations suggest that the PS sequences that confer the highest specificity and are most robust to solution conditions contain one or a few high affinity PSs and a stoichiometric or small excess of low affinity PSs. This observation is consistent with recent models [42], observation of multiple weak PSs in viral genomes [75], and recent *in vitro* measurements [95]. Our simulations identify multiple mechanisms by which PSs can confer specificity, depending on the protein sequence and solution conditions. Under conditions where protein subunit-subunit or sequence-independent protein-RNA interactions are too weak to nucleate assembly, PSs that enhance protein-RNA interactions and RNA-mediated protein-protein interactions can induce nucleation and facilitate subsequent assembly. Under conditions where strong subunit-subunit and nonspecific subunit-RNA interactions lead to multiple, geometrically incompatible partial capsids forming on individual RNAs, efficient and specific assembly can be realized by PS sequences that favor nucleation and rapid assembly of a single partial capsid. Finally, the simulations demonstrate that PSs can dramatically alter assembly pathways in comparison to non-cognate RNAs (Fig. 7), as observed in recent experiments [36], but that the effect on assembly pathways is sensitive to solution conditions.

Our simulation results suggest that rapid, specific assembly can proceed via a diverse ensemble of pathways, provided that RNA conformations can anneal during assembly through reversible interactions and/or cooperative RNA-protein rearrangements. This finding is consistent with the observation that proteins can fold by multiple, dissimilar pathways [101]. In particular, the RNA does not trace a Hamiltonian path, as has been inferred from structures of $T=3$ MS2 capsids [33] and assumed in other models [22, 42]. However, the expectation of a Hamiltonian path in those models was based on coupling between PS binding and subunit conformation (as occurs in the $T=3$ MS2 capsid), which is not present in our model of a $T=1$ capsid.

Dykeman et al. [22] recently showed that the gradually increasing protein concentration characteristic of an MS2-infected *E. coli* can increase specificity for a cognate RNA in comparison to assembly under a fixed protein concentration. Enhanced specificity arises in their model because during the initial stages of the reaction nucleation occurs only around cognate RNAs, similar to the behavior in our simulations for high salt and low ϵ_{ss} (Fig. 3). When the protein concentration increases during later stages it is rapidly consumed by growth of the partial capsid-cognate RNA complexes. While simulations with time-varying protein concentrations are beyond the scope of the present work, we anticipate that similar specificity enhancements would arise in our model.

Implications for experiments

Our simulations predict that the specificity conferred by PSs is sensitive to parameters that control the pathways and efficiency of sequence-independent assembly. While it has been previously suggested that the degree of specificity observed in *in vitro* experiments is sensitive to subunit concentration, the predicted phase diagrams reveal that varying ionic strength, *pH*, or protein-RNA binding sequences (through mutagenesis) could shift an experiment from selective to nonselective. This result may shed light on the varied degrees of specificity observed in previous competition experiments [18, 30, 31, 34, 35]. However, note that the location of boundaries within the predicted phase diagrams depend on the protein-RNA binding sequence[44]. Furthermore, experimentally measured specificity has been defined in different ways, depending on whether capsid protein is in excess or limiting. Our simulations find that these two conditions and definitions can lead to similar or different observed levels of specificity, depending on the location in parameter space (see section III A). An intriguing prediction from our simulations is that excess PSs (in comparison to the number of binding sites within a complete capsid) can increase assembly under some conditions by promoting exchange of improperly bound PSs. This prediction, as well as the general dependencies on the affinity and number of PSs, could be tested by constructing RNA fragments with varying numbers of high- and low-affinity PSs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge William Gelbart and Chuck Knobler for insightful discussions and critical reading of the manuscript, as well as Reidun Twarock for a critical reading of the manuscript and helpful discussions, particularly about the geometry of packaging site–capsid protein interactions. This work was supported by Award Number R01GM108021 from the National Institute Of General Medical Sciences. Computational resources were provided by the NSF through XSEDE computing resources (Maverick and Keeneland) and the Brandeis HPCC which is partially supported by the Brandeis Center for Bioinspired Soft Materials, an NSF MRSEC, DMR-1420382.

References

1. Perlmutter, Jason D.; Hagan, Michael F. Mechanisms of virus assembly. *Annu Rev Phys Chem.* 2015; 66(1):217–239. [PubMed: 25532951]

2. Routh, Andrew; Domitrovic, Tatiana; Johnson, John E. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci U S A.* 2012; 109(6): 1907–1912. [PubMed: 22308402]
3. Schneemann A. The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol.* 2006; 60:51–67. [PubMed: 16704342]
4. Bancroft, Jb; Hiebert, E.; Bracker, Ce. Effects Of Various Polyanions On Shell Formation Of Some Spherical Viruses. *Virology.* 1969; 39(4):924–930. [PubMed: 5358835]
5. Hohn, Thomas. Role of RNA in the assembly process of bacteriophage fr. *J Mol Biol.* 1969; 43(1): 191–200. [PubMed: 4897789]
6. Chen C, Kwak ES, Stein B, Kao CC, Dragnea B. Packaging of gold particles in viral capsids. *J Nanosci Nanotechnol.* 2005; 5(12):2029–2033.
7. Sun J, DuFort C, Daniel MC, Murali A, Chen C, Gopinath K, Stein B, De M, Rotello VM, Holzenburg A, Kao CC, Dragnea B. Core-controlled polymorphism in virus-like particles. *Proc Natl Acad Sci U S A.* 2007; 104(4):1354–1359. [PubMed: 17227841]
8. Hu, Yufang; Zandi, Roya; Anavitarte, Adriana; Knobler, Charles M.; Gelbart, William M. Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size. *Biophys J.* Feb 15; 2008 94(4):1428–1436. [PubMed: 17981893]
9. Sikkema FD, Comellas-Aragones M, Fokkink RG, Verduin BJM, Cornelissen Jlm, Nolte RJM. Monodisperse polymer-virus hybrid nanoparticles. *Org Biomol Chem.* 2007; 5(1):54–57. [PubMed: 17164905]
10. Brasch, Melanie; Cornelissen, Jeroen JLM. Relative size selection of a conjugated polyelectrolyte in virus-like protein structures. *Chem Commun.* 2012; 48(10):1446–1448.
11. Kostianinen, Mauri A.; Pietsch, Christian; Hoogenboom, Richard; Nolte, Roeland JM.; Cornelissen, Jeroen JLM. Temperature-Switchable Assembly of Supramolecular Virus-Polymer Complexes. *Adv Funct Mater.* 2011; 21(11):2012–2019.
12. Goicochea, Nancy L.; De, Mrinmoy; Rotello, Vincent M.; Mukhopadhyay, Suchetana; Dragnea, Bogdan. Core-like particles of an enveloped animal virus can self-assemble efficiently on artificial templates. *Nano Lett.* 2007; 7(8):2281–2290. [PubMed: 17645363]
13. Loo, LiNa; Guenther, Richard H.; Lommel, Steven A.; Franzen, Stefan. Encapsulation of nanoparticles by Red Clover Necrotic Mosaic Virus. *J Am Chem Soc.* Sep 12; 2007 129(36): 11111–11117. [PubMed: 17705477]
14. Kwak, Minseok; Minten, Inge J.; Anaya, Diana-Milena; Musser, Andrew J.; Brasch, Melanie; Nolte, Roeland JM.; Muellen, Klaus; Cornelissen, Jeroen JLM.; Herrmann, Andreas. Virus-like Particles Templated by DNA Micelles: A General Method for Loading Virus Nanocarriers. *J Am Chem Soc.* 2010; 132(23):7834–7835. [PubMed: 20481536]
15. Chang, Connie B.; Knobler, Charles M.; Gelbart, William M.; Mason, Thomas G. Curvature dependence of viral protein structures on encapsidated nanoemulsion droplets. *ACS Nano.* Feb; 2008 2(2):281–286. [PubMed: 19206628]
16. Malyutin, Andrey G.; Dragnea, Bogdan. Budding Pathway in the Templated Assembly of Viruslike Particles. *J Phys Chem B.* Sep.2013
17. Cheng, Fan; Tsvetkova, Irina B.; Khuong, Y-Lan; Moore, Alan W.; Arnold, Randy J.; Goicochea, Nancy L.; Dragnea, Bogdan; Mukhopadhyay, Suchetana. The Packaging of Different Cargo into Enveloped Viral Nanoparticles. *Mol Pharm.* 2013; 10(1):51–58. [PubMed: 22876758]
18. Comas-Garcia, Mauricio; Cadena-Nava, Ruben D.; Rao, ALN.; Knobler, Charles M.; Gelbart, William M. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *J Virol.* 2012; 86(22):12271–12282. [PubMed: 22951822]
19. Perlmutter, Jason D.; Qiao, Cong; Hagan, Michael F. Viral genome structures are optimal for capsid assembly. *eLife.* 2013; 2:e00632. [PubMed: 23795290]
20. Erdemci-Tandogan, Gonca; Wagner, Jef; van der Schoot, Paul; Podgornik, Rudolf; Zandi, Roya. RNA topology remodels electrostatic stabilization of viruses. *Phys Rev E.* Mar.2014 89:032707.
21. Bamunusinghe, Devinka; Seo, Jang-Kyun; Rao, ALN. Subcellular Localization and Rearrangement of Endoplasmic Reticulum by Brome Mosaic Virus Capsid Protein. *J Virol.* 2011; 85(6):2953–2963. [PubMed: 21209103]

22. Dykeman, Eric C.; Stockley, Peter G.; Twarock, Reidun. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc Natl Acad Sci U S A*. Apr 8; 2014 111(14): 5361–5366. [PubMed: 24706827]
23. Annamalai P, Rofail F, DeMason DA, Rao ALN. Replication-coupled packaging mechanism in positive-strand RNA viruses: Synchronized coexpression of functional multigenome RNA components of an animal and a plant virus in *Nicotiana benthamiana* cells by agroinfiltration. *J Virol*. 2008; 82(3):1484–1495. [PubMed: 18032497]
24. Kao CC, Ni P, Hema M, Huang XL, Dragnea B. The coat protein leads the way: an update on basic and applied studies with the Brome mosaic virus coat protein. *Mol Plant Pathol*. 2011; 12(4):403–412. [PubMed: 21453435]
25. Rao ALN. Genome packaging by spherical plant RNA viruses. *Annu Rev Phytopathol*. 2006; 44:61–87. [PubMed: 16480335]
26. Stockley, Peter G.; Twarock, Reidun; Bakker, Saskia E.; Barker, Amy M.; Borodavka, Alexander; Dykeman, Eric; Ford, Robert J.; Pearson, Arwen R.; Phillips, Simon EV.; Ranson, Neil A., et al. Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J Biol Phys*. 2013; 39(2):277–287. [PubMed: 23704797]
27. Pappalardo, Lucia; Kerwood, Deborah J.; Pelczer, Istvan; Borer, Philip N. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J Mol Biol*. 1998; 282(4):801–818. [PubMed: 9743628]
28. Lu K, Heng X, Summers MF. Structural Determinants and Mechanism of HIV-1 Genome Packaging. *J Mol Biol*. 2011; 410(4):609–633. [PubMed: 21762803]
29. D'Souza V, Summers MF. How retroviruses select their genomes. *Nat Rev Microbiol*. 2005; 3(8): 643–655. [PubMed: 16064056]
30. Sorger PK, Stockley PG, Harrison SC. Structure and Assembly of Turnip Crinkle Virus. 2. Mechanism of Reassembly In vitro. *J Mol Biol*. 1986; 191(4):639–658. [PubMed: 3806677]
31. Beckett D, Wu HN, Uhlenbeck OC. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J Mol Biol*. Dec; 1988 204(4):939–47. [PubMed: 3221401]
32. Bunka DHJ, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, Twarock R, Phillips SEV, Stockley PG. Degenerate RNA Packaging Signals in the Genome of Satellite Tobacco Necrosis Virus: Implications for the Assembly of a T=1 Capsid. *J Mol Biol*. 2011; 413(1):51–65. [PubMed: 21839093]
33. Dykeman EC, Grayson NE, Toropova K, Ranson NA, Stockley PG, Twarock R. Simple Rules for Efficient Assembly Predict the Layout of a Packaged Viral RNA. *J Mol Biol*. 2011; 408(3):399–407. [PubMed: 21354423]
34. Ling CM, Hung PP, Overby LR. Independent assembly of Qbeta and MS2 phages in doubly infected *Escherichia coli*. *Virology*. Apr; 1970 40(4):920–9. [PubMed: 4914647]
35. Zachary Porterfield J, Dhason Mary Savari, Loeb Daniel D, Nassal Michael, Stray Stephen J, Zlotnick Adam. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. *J Virol*. 2010; 84(14):7174–7184. [PubMed: 20427522]
36. Borodavka, Alexander; Tuma, Roman; Stockley, Peter G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc Natl Acad Sci U S A*. 2012; 109(39):15769–15774. [PubMed: 23019360]
37. Ford, Robert J.; Barker, Amy M.; Bakker, Saskia E.; Coutts, Robert H.; Ranson, Neil A.; Phillips, Simon EV.; Pearson, Arwen R.; Stockley, Peter G. Sequence-specific, RNA–protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *Journal of molecular biology*. 2013; 425(6):1050–1064. [PubMed: 23318955]
38. Bortz AB, Kalos MH, Lebowitz JL. New Algorithm for Monte Carlo Simulation of Ising Spin Systems. *J Comput Phys*. 1975; 17(1):10–18.
39. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem*. 1977; 81(25):2340–2361.
40. Jamalyaria F, Rohlfis R, Schwartz R. Queue-based method for efficient simulation of biological self-assembly systems. *J Comput Phys*. Mar 20; 2005 204(1):100–120.

41. Sweeney B, Zhang T, Schwartz R. Exploring the parameter space of complex self-assembly through virus capsid models. *Biophys J*. 2008; 94(3):772–783. [PubMed: 17921216]
42. Dykeman, Eric C.; Stockley, Peter G.; Twarock, Reidun. Building a viral capsid in the presence of genomic RNA. *Phys Rev E*. Feb 25.2013 87(2):022717.
43. Eigen M, Biebricher CK, Gebinoga M, Gardiner WC. The hypercycle coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry*. Nov 19; 1991 30(46):11005–11018. [PubMed: 1932025]
44. Perlmutter JD, Perkett MR, Hagan Michael F. Pathways for virus assembly around nucleic acids. *Journal of Molecular Biology*. 2014
45. Whitelam S, Jack RL. The Statistical Mechanics of Dynamic Pathways to Self-assembly. *Ann Rev Phys Chem*. 2015; 66:143–63. [PubMed: 25493714]
46. Hagan, Michael F. Modeling Viral Capsid Assembly. *Adv Chem Phys*. 2014; 155:1–68. [PubMed: 25663722]
47. Kler, Stanislav; Asor, Roi; Li, Chenglei; Ginsburg, Avi; Harries, Daniel; Oppenheim, Ariella; Zlotnick, Adam; Raviv, Uri. RNA Encapsulation by SV40-Derived Nanoparticles Follows a Rapid Two-State Mechanism. *J Am Chem Soc*. 2012; 134(21):8823–8830. [PubMed: 22329660]
48. Kler, Stanislav; Wang, Joseph Che-Yen; Dhason, Mary; Oppenheim, Ariella; Zlotnick, Adam. Scaffold properties are a key determinant of the size and shape of self-assembled virus-derived particles. *ACS Chem Biol*. Dec; 2013 8(12):2753–61. [PubMed: 24093474]
49. Wales DJ. The energy landscape as a unifying theme in molecular science. *Phil Trans R Soc A*. 2005; 363(1827):357–375. [PubMed: 15664888]
50. Ceres P, Zlotnick A. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry*. 2002; 41(39):11525–11531. [PubMed: 12269796]
51. Kegel WK, van der Schoot P. Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophys J*. 2004; 86(6):3905–3913. [PubMed: 15189887]
52. Fejer, Szilard N.; James, Tim R.; Hernandez-Rojas, Javier; Wales, David J. Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. *Phys Chem Chem Phys*. 2009; 11(12):2098–2104. [PubMed: 19280020]
53. Johnston IG, Louis AA, Doye JPK. Modelling the self-assembly of virus capsids. *J Phys : Condens Matter*. 2010; 22(10):104101. [PubMed: 21389435]
54. Schwartz R, Shor PW, Prevelige PE, Berger B. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys J*. 1998; 75(6):2626–2636. [PubMed: 9826587]
55. Hagan MF, Chandler D. Dynamic pathways for viral capsid assembly. *Biophys J*. 2006; 91(1):42–54. [PubMed: 16565055]
56. Hicks SD, Henley CL. Irreversible growth model for virus capsid assembly. *Phys Rev E*. 2006; 74(3):031912.
57. Nguyen HD, Reddy VS, Brooks CL. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett*. 2007; 7(2):338–344. [PubMed: 17297998]
58. Wilber AW, Doye JPK, Louis AA, Noya EG, Miller MA, Wong P. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J Chem Phys*. 2007; 127(8):085106. [PubMed: 17764305]
59. Nguyen HD, Brooks CL. Generalized structural polymorphism in self-assembled viral particles. *Nano Lett*. 2008; 8:4574. [PubMed: 19367856]
60. Nguyen, Hung D.; Reddy, Vijay S.; Brooks, Charles L. Invariant polymorphism in virus capsid assembly. *J Am Chem Soc*. 2009; 131(7):2606–14. [PubMed: 19199626]
61. Wilber AW, Doye JPK, Louis AA. Self-assembly of monodisperse clusters: Dependence on target geometry. *J Chem Phys*. 2009; 131(17)
62. Wilber, Alex W.; Doye, Jonathan PK.; Louis, Ard A.; Lewis, Anna CF. Monodisperse self-assembly in a model with protein-like interactions. *J Chem Phys*. 2009; 131(17):175102. [PubMed: 19895043]
63. Rapaport DC, Johnson JE, Skolnick J. Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation. *Comput Phys Commun*. 1999; 122:231–235.

64. Rapaport DC. Self-assembly of polyhedral shells: a molecular dynamics study. *Phys Rev E*. 2004; 70:051905.
65. Rapaport DC. The role of reversibility in viral capsid growth: A paradigm for self-assembly. *Phys Rev Lett*. 2008; 101:186101. [PubMed: 18999841]
66. Hagan MF. Controlling viral capsid assembly with templating. *Phys Rev E*. 2008; 77:051904.
67. Elrad OM, Hagan MF. Encapsulation of a polymer by an icosahedral virus. *Phys Biol*. 2010; 7:045003. [PubMed: 21149971]
68. Hagan MF, Elrad OM, Jack RL. Mechanisms of kinetic trapping in self-assembly and phase transformation. *J Chem Phys*. 2011; 135:104115. [PubMed: 21932884]
69. Mahalik JP, Muthukumar M. Langevin dynamics simulation of polymer-assisted virus-like assembly. *J Chem Phys*. 2012; 136(13):135101. [PubMed: 22482588]
70. Levandovsky, Artem; Zandi, Roya. Nonequilibrium assembly, retroviruses, and conical structures. *Phys Rev Lett*. 2009; 102(19):198102–198102. [PubMed: 19518999]
71. Archer, Eva J.; Simpson, Mark A.; Watts, Nicholas J.; O’Kane, Rory; Wang, Bangchen; Erie, Dorothy A.; McPherson, Alex; Weeks, Kevin M. Long-Range Architecture in a Viral RNA Genome. *Biochemistry*. 2013; 52(18):3182–3190. [PubMed: 23614526]
72. Athavale, Shreyas S.; Jared Gossett, J.; Bowman, Jessica C.; Hud, Nicholas V.; Williams, Loren Dean; Harvey, Stephen C. In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PloS one*. 2013; 8(1):e54384. [PubMed: 23349871]
73. Schroeder, Susan J.; Stone, Jonathan W.; Bleckley, Samuel; Gibbons, Theodore; Mathews, Deborah M. Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophys J*. Jul; 2011 101(1):167–75. [PubMed: 21723827]
74. Zeng Y, Larson SB, Heitsch CE, McPherson A, Harvey SC. A Model for the Structure of Satellite Tobacco Mosaic Virus. *J Struct Biol*. 2012; 180(1):110–116. [PubMed: 22750417]
75. Dykeman, Eric C.; Stockley, Peter G.; Twarock, Reidun. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J Mol Biol*. Sep; 2013 425(17):3235–49. [PubMed: 23763992]
76. Anderson, Joshua A.; Lorenz, Chris D.; Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J Comput Phys*. 2008; 227(10):5342–5359.
77. Nguyen, Trung Dac; Phillips, Carolyn L.; Anderson, Joshua A.; Glotzer, Sharon C. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. *Comput Phys Commun*. 2011; 182(11):2307–2313.
78. LeBard, David N.; Levine, Benjamin G.; Mertmann, Philipp; Barr, Stephen A.; Jusufi, Arben; Sanders, Samantha; Klein, Michael L.; Panagiotopoulos, Athanassios Z. Self-assembly of coarse-grained ionic surfactants accelerated by graphics processing units. *Soft Matter*. 2012; 8(8):2385–2397.
79. Widom B. Some topics in the theory of fluids. *J Chem Phys*. 1963; 39(11):2808–2812.
80. Kumar SK, Szleifer I, Panagiotopoulos AZ. Determination Of The Chemical-potentials Of Polymeric Systems From Monte-carlo Simulations. *Phys Rev Lett*. Jun 3; 1991 66(22):2935–2938. [PubMed: 10043657]
81. Deserno, Markus. Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications. *Macromol Rapid Commun*. 2009; 30(9–10):752–771. [PubMed: 21706661]
82. Hyeon, Changbong; Thirumalai, D. Chain length determines the folding rates of RNA. *Biophys J*. Feb; 2012 102(3):L11–3. [PubMed: 22325296]
83. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; 14(1):33–38. [PubMed: 8744570]
84. Ortega A, Amorós D, García de La Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic-and residue-level models. *Biophysical journal*. 2011; 101(4):892–898. [PubMed: 21843480]
85. Becker R, Döring W. *Ann Phys (Leipzig)*. 1935; 416:719.

86. Endres D, Zlotnick A. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J.* 2002; 83(2):1217–1230. [PubMed: 12124301]
87. Yoffe AM, Prinsen P, Gopal A, Knobler CM, Gelbart WM, Ben-Shaul A. Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci U S A.* 2008; 105(42):16153–16158. [PubMed: 18845685]
88. Elrad OM, Hagan MF. *Nano Lett.* 2008; 8(11):3850–3857. [PubMed: 18950240]
89. McPherson A. Micelle formation and crystallization as paradigms for virus assembly. *Bioessays.* 2005; 27(4):447–458. [PubMed: 15770675]
90. Devkota B, Petrov AS, Lemieux S, Boz MB, Tang L, Schneemann A, Johnson JE, Harvey SC. Structural and Electrostatic Characterization of Pariacoto Virus: Implications for Viral Assembly. *Biopolymers.* 2009; 91(7):530–538. [PubMed: 19226622]
91. Hagan MF. A theory for viral capsid assembly around electrostatic cores. *J Chem Phys.* 2009; 130:114902. [PubMed: 19317561]
92. Cadena-Nava, Ruben D.; Comas-Garcia, Mauricio; Garmann, Rees F.; Rao, ALN.; Knobler, Charles M.; Gelbart, William M. Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *J Virol.* 2012; 86(6): 3318–3326. [PubMed: 22205731]
93. Garmann, Rees F.; Comas-Garcia, Mauricio; Gopal, Ajaykumar; Knobler, Charles M.; Gelbart, William M. The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. *J Mol Biol.* Mar 6; 2014 426(5 SI):1050–1060. [PubMed: 24148696]
94. Comas-Garcia, Mauricio; Garmann, Rees F.; Singaram, Surendra W.; Ben-Shaul, Avinoam; Knobler, Charles M.; Gelbart, William M. Characterization of Viral Capsid Protein Self-Assembly around Short Single-Stranded RNA. *J Phys Chem B.* 2014
95. Patel, Nikesh; Dykeman, Eric C.; Coutts, Robert HA.; Lomonosoff, George P.; Rowlands, David J.; Phillips, Simon EV.; Ranson, Neil; Twarock, Reidun; Tuma, Roman; Stockley, Peter G. Revealing the density of encoded functions in a viral RNA. *Proc Natl Acad Sci USA.* Feb.2015
96. Zlotnick A. Are weak protein-protein interactions the general rule in capsid assembly? *Virology.* 2003; 315:269–274. [PubMed: 14585329]
97. Jack RL, Hagan MF, Chandler D. Fluctuation-dissipation ratios in the dynamics of self-assembly. *Phys Rev E.* 2007; 76:021119.
98. Grant, James; Jack, Robert L.; Whitlam, Stephen. Analyzing mechanisms and microscopic reversibility of self-assembly. *J Chem Phys.* 2011; 135(21)
99. Kivenson A, Hagan MF. Mechanisms of viral capsid assembly around a polymer. *Biophys J.* 2010; 99(2):619–628. [PubMed: 20643082]
100. Hu, Tao; Shklovskii, BI. Kinetics of viral self-assembly: Role of the single-stranded RNA antenna. *Phys Rev E.* 2007; 75:051901.
101. Bowman, Gregory R.; Pande, Vijay S. Protein folded states are kinetic hubs. *Proceedings of the National Academy of Sciences.* 2010; 107(24):10890–10895.
102. Zlotnick, Adam; Zachary Porterfield, J.; Wang, Joseph Che-Yen. To Build a Virus on a Nucleic Acid Substrate. *Biophys J.* 2013; 104(7):1595–1604. [PubMed: 23561536]
103. Hu T, Zhang R, Shklovskii BI. Electrostatic theory of viral self-assembly. *Physica A.* 2008; 387(12):3059–3064.
104. Belyi VA, Muthukumar M. Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci U S A.* 2006; 103(46):17174–17178. [PubMed: 17090672]

Research Highlights

- During the life cycle of many viruses, a protein capsid must selectively assemble around the genomic RNA.
- RNA encapsidation is driven by nonspecific electrostatics and protein interactions with specific RNA sequences known as packaging sites.
- We use dynamical computer simulations to study how the selectivity conferred by packaging sites depends on solution conditions, protein interactions, and other control parameters.
- The simulations demonstrate that packaging sites enable extreme specificity, but only in certain regions of parameter space.
- Assembly around cognate (with packaging sites) and non-cognate (no packaging sites) RNAs can proceed by different classes of pathways.
- Understanding how nonspecific and specific interactions contribute to the efficiency and selectivity of assembly may promote efforts to develop antiviral agents that interfere selective encapsidation of viral genomes.

FIG. 1.

(A) Model schematic showing components responsible for subunit-subunit interactions: subunits are bound together by attractor pseudoatoms ('A'), and the Top ('T') and Bottom ('B') pseudoatoms guide the subunits towards the correct geometry (see SI). (B) Schematic with components responsible for attractive interaction with the RNA (drawn in red) and packaging site ('PS'): positively charged ARM ('+') and PS Receptor ('PSR'). The 'Excluder' pseudoatoms, which represent the excluded volume of the capsid shell, are located within the black pentagons; to aid visibility, they are not explicitly drawn here. Snapshots here and throughout the article are colored as follows: blue=excluders, green=attractors, yellow=ARM, red=RNA, orange=PS.

FIG. 2.

The effect of PSs and solution conditions on assembly yields and products. **(A,B)** The most prevalent assembly product is shown as a function of ionic strength (C_{salt}) and subunit-subunit attraction well-depth (ϵ_{ss}) for assembly around **(A)** a non-cognate RNA (polyelectrolyte without PS), and **(B)** a cognate RNA with 1 high affinity (HA, $\epsilon_{PS} = 20K_B T$) PS and 25 low affinity (LA, $\epsilon_{PS} = 5K_B T$) PSs. A legend showing the outcome and a representative simulation snapshot corresponding to each symbol is presented in **(C)**. **(D,E)** The yield of well-formed capsids assembled around **(D)** the non-cognate RNA or **(E)** the cognate RNA with the PS sequence as in **(B)**. In each simulation the RNA length corresponds to the thermodynamic optimal length for the non-cognate at the simulated value of C_{salt} , and ranges from 350 to 575 RNA segments (see Fig. S7).

FIG. 3.

Selectivity for RNA containing 1 HA PS + 25 LA PS competing against a non-cognate RNA at equal concentrations $r_{ex} = 1$, (**A**) estimated from the data in Fig. 2 using Eq. 1 and (**B**) measured in direct competition simulations. (**C**) Selectivity for RNA containing 1 HA PS + 25 LA PS competing against excess non-cognate RNA, $r_{ex} = 10$. As in Fig 2, in each simulation the optimal RNA length is used based on the results in Fig. S7. In the explicit competition simulations of (**B**) the concentration of subunits is the same as used in the assembly simulations (Fig. 2). Error bars indicate 95% confidence intervals calculated using bootstrapping.

FIG. 4. Yield as a function of number of PS, N_{PS} , at low (**A**) and high (**B**) salt concentrations. Note that for these parameters yield is zero in the absence of PSs. PSs are either all low affinity (LA) (■), all high affinity (HA) (● symbols), or the Combo sequence with 1 HA and N_{PS} LA PSs (▲ symbols). For these cases, the HA PS is placed in the center of the RNA. Results from sets of simulations with the HA PS placed in the terminal position are shown as ○ symbols. The result from simulations with the PS binding site placed in the center of the subunits is shown as a ♦ symbol. Note that there are 20 PS binding sites in a complete capsid, so $N_{PS} = 20$ is the stoichiometric value. Snapshots illustrate the trend in dominant outcomes with increasing PS number. Error bars indicate 95% confidence intervals calculated using bootstrapping.

FIG. 5.

The assembly pathway order parameter n_{free} measured from simulations for (left) the non-cognate RNA and (center) the cognate RNA with the Combo PS sequence, $N_{\text{PS}} = 25$. (Right) The change in n_{free} due to PSs.

FIG. 6.

Snapshots from typical assembly trajectories without and with PSs (the cognate RNA here is the combo sequence with 1 HA and 25 LA PSs) for low and high salt concentrations. PSs are depicted as orange spheres with exaggerated size to improve visibility.

FIG. 7. (A–F) Radius of hydration R_H as a function of simulation time steps for assembly trajectories performed at indicated parameter values, for non-cognate RNA (■ symbols) and cognate RNA (● symbols). The R_H values before subunits are introduced are shown as ▲ symbols. The subunit-subunit interaction energy ϵ_{ss} increases from left to right. In the top row (A–C), the subunit ARM charge is (+5), and the RNA length is 575 segments; in the second row (D–F), the subunit ARM charge is (+10), and the RNA length is 910 segments. (G) Snapshots from simulations corresponding to panel (E), with non-cognate RNA on the left and cognate RNA on the right.

FIG. 8.

(A) Path duration during a dynamic trajectory for RNAs within a preassembled capsid. (B) Schematic representation of RNA path within the capsid at intervals of 5×10^7 timesteps. The line indicates the RNA path, with line color and width changing gradually with contour length for clarity. (C) Snapshots and schematics indicating non-optimal RNA paths which lead to stalled assemblies. PSs are shown with exaggerated size to improve visibility. Segments of interest are shown in green.

TABLE I

Table comparing parameter values for the model system and biological or experimental *in vitro* values.

Parameter	Model Value(s)	Biological Reference
Capsid protein concentration	12 μM subunits	1–10 μM <i>in vitro</i> capsid protein [37]
Capsid radius	7.3 nm	7.2 nm STMV, 7.7 nm STNV
Subunit-subunit binding affinity (K_D)	45 μM –7M	>9 mM CCMV, ~3 mM HBV, $\ll$ 1 mM SV40 [102]
Subunit-subunit binding affinity (kcal/mol)	–6 to 2 kcal/mol	>–2.8 CCMV, ~–3.4 HBV, $\ll$ –4.1 kcal/mol SV40 [102]
LA Packaging site affinity (K_D)	1.3 μM	<1.5 μM MS2 [75]
Number of packaging sites, N_{PS}	1–40	*30,35 STNV [32], **30,33 STMV [74]
ARM charge (number of positive charges)	5,10	6–31 [103, 104]
Salt concentrations, C_{salt}	1–500 mM	150 mM <i>in vivo</i> , 100 mM [93] & 250 mM [47] <i>in vitro</i>
Non-cognate/Cognate competition ratio, r_{ex}	1–10	300 <i>in vivo</i> [22], 1–50 <i>in vitro</i> [35]

* Proposed number of PSs for two STNV strains [32].

** Proposed numbers of PSs in the STMV genome based on folding and crystallographic data [73, 74].