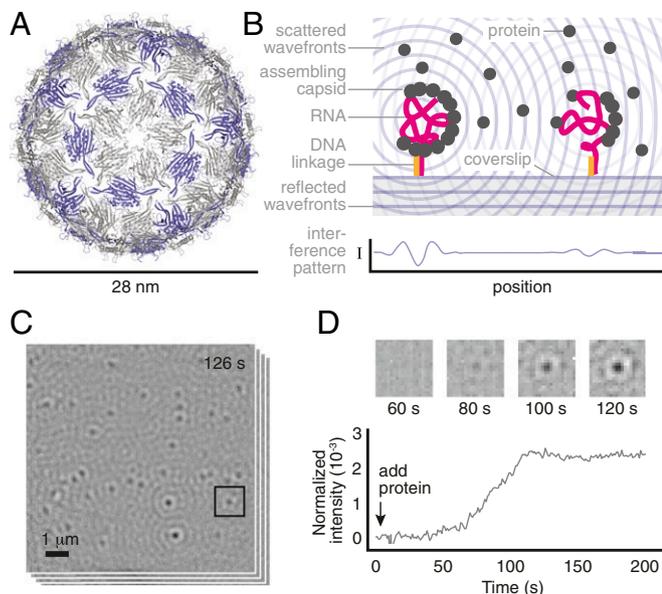Fig. 1. Overview of the measurement. (*A*) A structural model of the MS2 capsid (PDB ID: 2ms2) shows its small size and $T = 3$ structure. The 2 coat-protein dimer configurations are shown in gray and purple. (*B*) We inject a solution of unassembled dimers over a coverslip on which MS2 RNA strands are tethered by DNA linkages. As dimers bind to the RNA, the resulting particles scatter light. The particles appear as dark, diffraction-limited spots because of destructive interference between the scattered light and a reference beam. (*C*) We monitor many such spots in parallel. Shown is a typical image of the field of view, taken 126 s after adding 2-μM dimers and representing an average of 1,000 frames taken at 1,000 frames per second. (*D*) The intensity of each spot is proportional to the number of bound proteins within each particle and changes in the intensity as a function of time reveal the assembly kinetics of each particle. The darker the spot is, the larger its intensity. (*D*, *Top*) Time series of images for the boxed spot in *C*. (*D*, *Bottom*) Intensity trace for the same spot using a 1,000-frame average. We discuss the relationship between intensity and number of bound proteins, as well as how we calculate the spot intensity, in *Materials and Methods* and *SI Appendix*.

kinetics of individual particles as well as the kinetics of the ensemble. The ensemble is characterized by traces that persist at a low intensity before rising rapidly and then plateauing at higher intensity (Fig. 2*A* and *SI Appendix*, Fig. S1). When we inject 2-μM coat-protein dimers, we find that the majority of traces (about 85%) plateau at a final intensity consistent with or slightly smaller than that of a complete wild-type capsid (*SI Appendix*, Fig. S2), with the remainder plateauing at significantly higher intensities. By contrast, in the absence of RNA, there are few traces, and they increase slowly and continuously throughout the experiment (*SI Appendix*, Fig. S3).

In a separate control experiment, we use transmission electron microscopy (TEM) to image the structures that assemble around surface-tethered RNA. The images reveal that most of the assembled structures are proper capsids, with a few partial capsids and larger structures visible (Fig. 2*B* and *SI Appendix*, Fig. S4). We therefore infer that capsids can indeed assemble around tethered RNA strands and that traces that reach intensities similar to those of wild-type capsids represent the formation of complete or nearly complete capsids.

With this understanding, we examine what the variations among individual traces reveal about the assembly pathways. A key observation is that assembly is not synchronous: The "start time," the time at which the intensity rapidly increases, varies from particle to particle (Fig. 2*A* and *SI Appendix*, Fig. S1). We find that the cumulative distribution of start times t is fitted well by an exponential function $A\left(1 - \exp\left[-(t - t_0)/\tau\right]\right)$ (Fig. 2*C*),

where A is the plateau value, t_0 is the delay before the start time of the first particle, and τ is the characteristic time. The values and uncertainties of the fit parameters, as estimated from fits to repeated measurements, are $A = 57 \pm 7$, $t_0 = 60 \pm 20$ s, and $\tau = 100 \pm 20$ s (SD) (*SI Appendix*, Fig. S5).

The delay t_0 before the first start time likely results from diffusion and a concentration threshold for assembly. We know such a threshold exists because we see no assembly when we inject 1-μM coat-protein dimers (*SI Appendix*, Fig. S6). We therefore expect assembly to be delayed until the dimer concentration at the coverslip surface reaches the threshold. Since the threshold is between 1 and 2 μM, we can estimate t_0 from the characteristic time for dimers to diffuse from the 2-μM injected fluid to the surface. Our estimate, 30 to 55 s (*SI Appendix*, Fig. S7), is close to the observed t_0 of 60 ± 20 s.

The broad distribution of start times, however, does not appear to result from diffusion. We would expect diffusion-limited growth only if the protein concentration around each RNA were to vary across the 10-μm field of view. But the time for a dimer to diffuse 10 μm is only 1 s, much smaller than the width of the distribution. Furthermore, after the initial delay, we estimate that about 1,000 coat-protein dimers are within 1 μm of each RNA. At this concentration, the pool of coat proteins is not significantly depleted by assembly, and fluctuations in concentration are negligible. We conclude that the observed intensity traces do not result from variations in protein concentration.

Taken together, these findings rule out assembly pathways that begin with diffusion-limited aggregation and point strongly to those involving nucleation—a process in which an initially
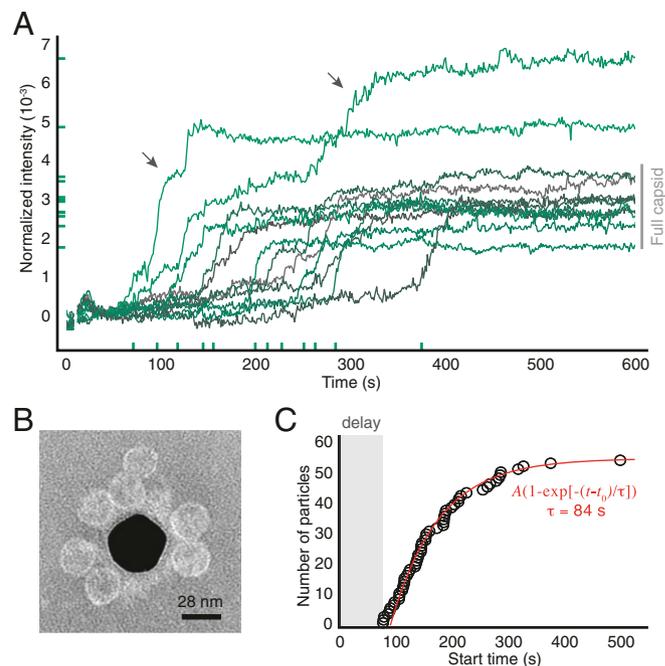
Fig. 2. Assembly of 2-μM coat-protein dimers around surface-tethered RNA strands. (*A*) Intensity traces for 12 randomly chosen particles from one experiment. *x*-axis ticks show the start times and *y*-axis ticks show the final intensities. Gray bar indicates the intensity range corresponding to wild-type capsids. Arrows show 2 traces corresponding to overgrown particles. (*B*) Negatively stained TEM image of particles assembled around RNA strands tethered to a gold nanoparticle (dark region at center). We use a nanoparticle as the substrate because TEM cannot image through a coverslip. (*C*) The cumulative distribution of start times of all of the traces in the experiment is well fitted by an exponential with delay time t_0 of 92 s and a characteristic time τ of 84 s (see *SI Appendix* for fit results from repeated experiments). Uncertainties in the start times are smaller than the diameter of the circles.

Garmann et al.

www.manaraa.com

unstable cluster of proteins called a "nucleus" must grow larger than a critical size before growth proceeds (17). Below this size, a growing nucleus becomes increasingly unstable, such that forming a critical nucleus corresponds to overcoming a free-energy barrier (18). The broad distribution of start times that we measure is evidence of such a barrier and is consistent with a nucleated process. That the shape of the distribution is well fitted by a single exponential suggests that there is a single, well-defined nucleation barrier, with a nucleation time of τ (see *SI Appendix* for how we rule out other barriers associated with possible nonspecific interactions between the RNA and the surface).

Fluctuations in the intensity reveal further information about the nucleation event. Before the start time, the fluctuations are consistent with those expected from shot noise, which, as noted above, corresponds to 6 dimers at 1-s averaging. This measurement indirectly constrains the critical nucleus size: If we assume that proteins bind only transiently to the RNA prior to nucleation, then subcritical nuclei larger than 6 dimers do not survive for longer than 1 s. However, if small numbers of proteins bind permanently to the RNA, then these proteins might be obscured by long-time drift in the measurement.

Owing to the small apparent size of the critical nucleus, most of the dimers in a capsid must add during the subsequent growth phase. We characterize the growth by examining the rise in intensity of each trace after nucleation. We find that the "growth time," the time required for a particle to reach the size of a complete capsid after it nucleates, varies from particle to particle, ranging from 30 to over 200 s (Fig. 2*A* and *SI Appendix*, Fig. S1). Of the particles that grow into a complete capsid, some grow at a constant rate, others slow as they approach completion, and still others contain intermediate pauses lasting up to 25 s (Fig. 2*A* and *SI Appendix*, Fig. S1). Despite these differences, essentially all of the traces are monotonic, with little or no observable disassembly steps.

These observations provide quantitative information about the kinetics of capsid growth and also reveal qualitative features of the growth pathways. The growth times are 4 orders of magnitude larger than the timescale required for a capsid's worth of dimers to collide with the RNA by diffusion (*SI Appendix*), suggesting that only a small fraction of these collisions leads to growth. And while the broad distribution of growth times and the varying shapes of the traces suggest many different growth pathways, the absence of any observable disassembly suggests that the system avoids pathways that involve bulk shedding of bound proteins.

Furthermore, we can glean information about the failure modes of the assembly process from traces of particles that grow significantly larger than a capsid. Because different assembly pathways can fail in different ways, resolving these failure modes provides valuable information about the operative pathways. We find that most traces of overgrown particles pause at an intensity consistent with that of a capsid before rising again to a higher plateau value (Fig. 2*A* and *SI Appendix*, Fig. S1). TEM images of particles assembled around untethered RNA show that most overgrown structures consist of a capsid attached to a second partial or complete capsid (*SI Appendix*, Fig. S8). These observations suggest that overgrown particles result from a second nucleation event that occurs before the growing capsid has fully packaged the RNA.

To test this hypothesis, we vary the assembly kinetics by adjusting the concentration of coat protein (Fig. 3*A* and *SI Appendix*, Figs. S9 and S10). We find that the nucleation time decreases with increasing protein concentration, from 160 ± 40 s at 1.5 μM of dimers to 11 ± 5 s at 4 μM (SD) (Fig. 3*B* and *SI Appendix*, Fig. S5). This decrease is accompanied by an increase in the fraction of overgrown particles, from about 5% at 1.5 μM dimers to over 40% at 4 μM (Fig. 3*C*). TEM images of untethered parti-
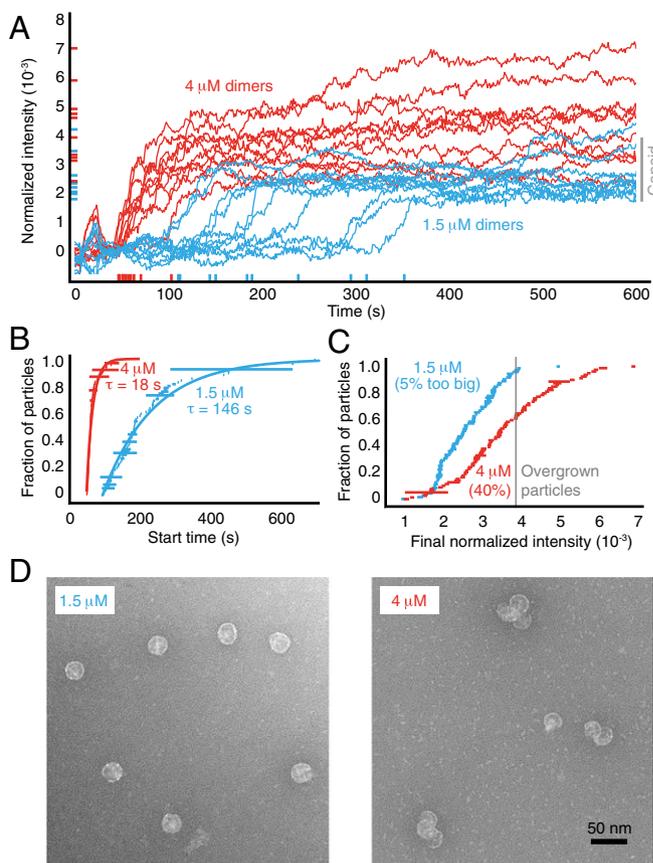
Fig. 3. Assembly kinetics at different protein concentrations. (*A*) Intensity traces for 10 randomly chosen particles at 1.5-μM and 4-μM coat-protein dimers. (*B*) Cumulative distributions of the start times show that the rate of nucleation increases with protein concentration. The data are fitted by an exponential with a characteristic nucleation time τ, as described above. The length of each horizontal bar represents the uncertainty in each time measurement. (*C*) Cumulative distributions of the final intensities show that the fraction of overgrown particles increases with protein concentration. The length of each horizontal bar is the SD calculated from the last 50 s of each trace. (*D*) TEM images of particles assembled around untethered RNA. At 1.5 μM protein (*Left*), most particles appear to be capsids. At 4 μM protein (*Right*), many particles are clusters of partial capsids.

cles reveal sizes that roughly correspond to the final intensities seen in the traces, with many of the overgrown particles consisting of bunches of partial or nearly complete capsids (Fig. 3*D* and *SI Appendix*, Fig. S8).

The growth times also decrease with increasing protein concentration, but less rapidly than do the nucleation times (Fig. 4*A*). Thus, we conclude that the RNA, which promotes nucleation at low protein concentration, creates a competition between nucleation and growth that can lead to overgrown structures at higher concentration, as sketched in Fig. 4*B*. This pathway provides a possible explanation for the "monster" (5) and "multiplet" (19) structures observed with other RNA viruses.

Discussion

Our measurements on individual MS2 capsids enable us to infer features of the assembly pathways that are obscured in bulk measurements, such as the presence of a nucleation barrier or the lack of substantial disassembly steps. We do not claim that these features must be present in assembly in vivo; indeed, we are able to make such inferences only because we perform our measurements outside the crowded environment of the cell. Furthermore, these features may differ for other viruses and other
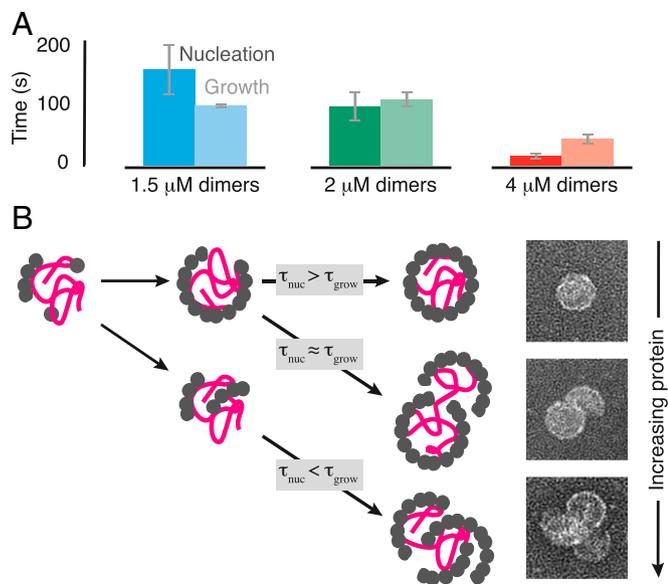
Fig. 4. Relative timescales of nucleation and growth and the inferred assembly pathways. (*A*) Measured nucleation times (τ_{nuc}) and median growth times (τ_{grow}) at different protein concentrations. Error bars represent the SD from 3 replicate experiments. (*B*) Cartoon of the inferred assembly pathways. At low protein concentration, τ_{nuc} is large compared to τ_{grow}, and a nucleus of coat proteins forms on the RNA and then grows into a proper capsid. At moderate protein concentration, τ_{nuc} is comparable to τ_{grow}, and a second nucleus can form on the RNA before the first one has finished growing, leading to an overgrown structure consisting of a nearly complete capsid attached to a partial capsid. At high protein concentration, τ_{nuc} is smaller than τ_{grow}, and multiple nuclei can form and grow, leading to an overgrown structure consisting of many partial capsids. Example TEM images of the endpoints of each pathway are shown at *Right*.

experimental conditions. Nonetheless, in vitro measurements performed using a range of viruses and experimental conditions have proved critical in developing an understanding of the physics of virus self-assembly (20).

Whether in vitro or in vivo, the coat proteins and RNA strands of many $T = 3$ viruses are able to pull off a remarkable feat: spontaneous self-assembly into a nontrivial structure with high yield. That this process is difficult to replicate in synthetic systems shows that the physics is still not well understood. We therefore examine our results in the context of general, physical models of virus self-assembly. There is a plethora of such models, which differ dramatically in terms of the role played by the RNA. By comparing our experimental results with the predictions of such models, we aim to determine what physical mechanisms are consistent with our data and what can be done to arrive at an assembly model that quantitatively agrees with experiment.

We first note that our results caution against the construction of assembly models based primarily on static measurements, such as structural data or equilibrium binding affinities. Such models suggest that assembly follows deterministically from the formation of specific protein–protein contacts (13) or high-affinity protein–RNA interactions (21). In contrast, the complex assembly pathways and failure modes reported here call for models that capture highly dynamic and collective processes.

Nucleation-and-growth models account for the collective nature of assembly. In nucleation models of empty (22) and protein-scaffolded capsids (23), which assemble without RNA, the critical nucleus forms in bulk. But a recent computational model that includes RNA as a homogeneous linear polymer

(24) explores 2 additional nucleation mechanisms. In one, the nucleus forms from a small cluster of proteins adsorbed to a bare RNA strand. In the other, it forms from proteins that have both adsorbed and aggregated along the entire strand.

Our observations suggest that nucleation happens directly on the RNA strand, rather than in bulk or in an aggregate. That the assembly is nucleated is evidenced by the broad distribution of times preceding assembly, as discussed in *Results*. If nucleation occurred in bulk, we would expect large numbers of empty capsids, but gel electrophoresis shows that the assembly products contain RNA (*SI Appendix*, Fig. S11). If the nucleation occurred in a disordered aggregate of proteins on the RNA, we would expect to see the intensity traces rise above the noise level during the aggregation process, but we do not see any such increase. Thus, we conclude that nucleation occurs on parts of the RNA that are bare or nearly bare.

While the diffraction limit precludes us from directly measuring where the nucleus forms on the RNA strand, our results do allow us to examine a longstanding assumption about the sequences involved. In MS2, coat proteins bind with high affinity (K_d of order 1 nM) to a 19-base sequence of the RNA termed the "translational operator" (10) and with much lower affinities (10 to 1,000 nM) to other 19-base sequences, suggesting that this operator acts as the nucleation site (25). However, our observation of a micromolar concentration threshold for assembly suggests that nucleation is not dominated by nanomolar-affinity binding between coat proteins and the translational operator, but likely involves other, weaker interactions. That a previous study reported assembly at 10-fold smaller protein concentrations (12) is not at odds with our observed threshold. Instead, the difference may reflect the different assembly buffer used, as described in *SI Appendix*.

Similarly, while we cannot determine the height of the nucleation barrier or size of the critical nucleus directly from our measurements, we can place a limit on the size of the nucleus from the intensity fluctuations, as discussed in *Results*. Owing to noise in our measurement and the finite averaging time, this limit is a combination of size and lifetime. Specifically, we find that subcritical clusters must be either very small (fewer than 6 dimers) or very short-lived (less than our averaging time, which is 1 s). If small numbers of proteins bind permanently to the RNA, they may also contribute to forming the nucleus. In either case, these size and lifetime bounds provide specific, quantitative constraints on models of assembly.

The apparent small size of the critical nucleus raises the question of how the growth process ensures high yields. In simplified models of $T = 1$ capsids, where all of the protein configurations are identical (26), the formation of a nucleus may be sufficient to ensure that growth leads to the correct structure. This need not be the case for $T = 3$ capsids, where there are inequivalent protein configurations and, consequently, many more possible misassembled structures, including those with the wrong local curvature (27). Our observation of monotonic growth following nucleation suggests that incoming proteins readily adopt configurations within the growing capsid that ensure the proper curvature. Even at high protein concentration, where the misassembled particles consist of multiple partial or full capsids, the curvature of these capsids does not appear to deviate from that of single, well-formed capsids.

One possible explanation is that interactions between the assembling proteins bias the pathways toward proper capsids. A recent model based on continuum elasticity theory (28) proposes that coat proteins adopt the correct configurations in the capsid as it grows in order to minimize the elastic stress. This model predicts that the number of proteins in the growing capsid should increase monotonically with time, a prediction that is consistent with our measurements of the growth phase. To further test this model, mutant coat proteins with different interactions might be

used to probe the role of elastic stress in controlling the growth pathways.

Another explanation is that protein–RNA interactions direct the growth pathways. Some models suggest that capsid growth is dominated by general, electrostatic interactions between the coat proteins and the RNA (29). However, a more recent model proposes that "multiple dispersed, specific interactions" (ref. 7, p. 74) between proteins and bits of the RNA secondary structure guide the incoming proteins into the correct local configurations. Recent experiments with MS2—including biochemical (30), structural (31), and bulk kinetic (12) measurements—provide indirect evidence that such structures are involved in capsid assembly but do not resolve how they affect the growth process. Because our method is compatible with RNA strands of arbitrary sequence and assembly buffers of arbitrary ionic strength, future studies can test these models by modifying the individual RNA structures or the strength of the electrostatic interactions and then measuring changes in the growth pathways.

Conclusions

Our measurements of the nucleation threshold, nucleation time, subcritical fluctuations, and growth time in MS2 provide important quantitative constraints on simulations of the assembly pathways (24). As a result, the structures of the critical nucleus and subsequent intermediate states, which have long eluded direct imaging methods, might now be inferred through quantitative comparisons of our measurements and such simulations. Understanding the assembly pathways for viral capsids in this level of detail may inform strategies for blocking the assembly of pathogenic viruses (32) or for engineering synthetic capsids (33).

Finally, while our observations are specific to the assembly pathways in vitro, they supply a physical hypothesis for viral replication in vivo. Because RNA inside a capsid cannot be translated or replicated, RNA viruses must delay capsid assembly and genome packaging until many copies of their proteins and genome are produced. An assembly pathway with a concentration threshold for nucleation—such as we observe—would provide such a delay. Moreover, this delay would allow the concentration of newly replicated viral RNA to increase relative to the background concentration of host RNA strands, increasing the likelihood that the viral RNA is packaged.

Materials and Methods

Growth of MS2 and Purification of Its Coat Protein and RNA. We grow wild-type MS2 (a gift from Peter Stockley at the University of Leeds, Leeds, UK) by infecting liquid cultures of *E. coli* strain C3000 (also a gift from Peter Stockley) and purifying the progeny viruses following the protocols of Strauss and Sinsheimer (34). We purify coat-protein dimers (2 × 13.7 = 27.4 kDa) from the virus particles following the cold acetic acid method described by Sugiyama, Hebert, and Hartman (2). We purify MS2 RNA (1.1 MDa) from freshly grown virus particles using an RNA extraction kit (RNeasy; Qiagen). Our procedure for assessing the concentration and purity of these materials is described in *SI Appendix*, Figs. S11 and S12.

Microscope. Our microscope is similar to that described by Ortega-Arroyo et al. (35) and is configured in wide-field mode without beam scanning. Details are given in *SI Appendix*, Fig. S13. We use 450-nm-wavelength light with an illumination intensity set to ∼3 kW/cm². Control experiments at different illumination intensities show that the intensity we use does not affect the assembly kinetics (*SI Appendix*, Fig. S14). The total field of view is 140 pixels × 140 pixels (9.8 μm × 9.8 μm). Images are recorded at 1,000 Hz. To combat mechanical drift over measurement durations of 900 s, the position of the coverslip relative to the objective is actively stabilized to a few nanometers in all 3 dimensions using piezoelectric actuators, as detailed in *SI Appendix*.

Image Intensity. Assembling particles appear as dark, diffraction-limited spots. The intensity of each spot is approximately linearly proportional to the number of proteins bound to the RNA strand. The intensity of a spot is $I = I_r + I_s + 2\sqrt{I_r I_s} \cos \phi_{rs}$, where I_r is the intensity of the reflected wave, I_s the intensity of the scattered wave, and ϕ_{rs} the phase difference between the two. The term I_s can be neglected since the scattered light is dim compared to the reflected light, so the normalized intensity $I_{norm} = I/I_r - 1$ is proportional to the total polarizability of the assembling particle (15), which is approximately the sum of a protein component and an RNA component. We previously showed that this linear superposition is a good approximation for bacteriophage λ particles, which consist of a protein capsid surrounding a single molecule of densely packed DNA (36). Here, because the RNA component is static, it is part of the background and subtracted away by our image-processing routine, which is described in *SI Appendix*. As a result, the normalized intensity is linearly proportional to the number of proteins in the assembling particle, consistent with other studies of multiprotein complexes (15).

Microscope Calibration. We calibrate the intensity measurements from the microscope using 2 particles of known mass: MS2 RNA strands and wild-type MS2 virus particles (*SI Appendix*, Fig. S2). We infer the intensity distribution corresponding to wild-type capsids by convolving the intensity distribution of the wild-type particles with the negative of the intensity distribution of the RNA. We use this inferred intensity distribution to estimate the intensity of full capsids that assemble in our experiments, as shown by the gray bars in Figs. 2A and 3A. The distribution is much broader than the expected mass distribution for wild-type capsids, which are quite monodisperse. Thus, the breadth of the distribution reflects the uncertainty of our measurement.

Coverslip Functionalization and RNA Binding. We adapt the protocols described by Joo and Ha (37) to coat glass coverslips with a layer of polyethylene glycol (PEG) molecules, about 1% of which are functionalized with a 20-base DNA strand. The sequence of the surface-bound DNA is GGTTGGTTGGTTGGTTGGTT. We further decorate the coverslips with PEG-passivated 30-nm gold particles that serve as tracer particles for active stabilization. Immediately before each experiment, we incubate the coverslips with a detergent solution containing 0.2% Tween-20 to passivate any bare patches in the PEG coating. Then we tether MS2 RNA strands to the coverslips using a 60-base DNA linker strand: The 40 bases at the 5′ end of the linker are complementary to the 40 bases at the 5′ end of the RNA, and the remaining 20 bases are complementary to the sequence of the surface-bound DNA (*SI Appendix*, Fig. S15). The sequence of the linker is CGACAGGAAGTTGAGCAGGACCCCGAAAGGGGTCCCACCCAAC-CAACCAACCAACCAACC. We hybridize the RNA to the linker by thermal annealing in hybridization buffer (50 mM Tris·HCl, pH 7.0; 200 mM NaCl; 1 mM EDTA) and then allow the resulting RNA–DNA complexes to bind to the DNA-functionalized surface at room temperature. Surface binding is highly specific (*SI Appendix*, Fig. S16). Details of the functionalization, thermal annealing, and surface binding steps are given in *SI Appendix*.

Assembly Kinetics Experiment. We introduce the coat-protein dimers to the surface-tethered RNA strands using home-built flow cells. The roof and sides of the cells are made of acrylic, and the floor consists of the RNA-functionalized coverslip described above. Details for constructing the flow cells are given in *SI Appendix*, Fig. S17. To perform the assembly experiment, we flush the flow cell with assembly buffer (42 mM Tris·HCl, pH 7.5; 84 mM NaCl; 3 mM acetic acid; 1 mM EDTA), start recording a movie using the interferometric scattering microscope, and then inject 10 μL of a specified concentration of unassembled dimers in assembly buffer using a syringe pump. The injected volume is roughly 3 times that of the sample chamber. The injection starts 4 s into the movie and continues for 20 s. An estimate of how long it takes for the injected dimers to reach the surface-tethered RNA strands is described in *SI Appendix*, Fig. S7. We perform 3 replicate experiments for all interferometric scattering microscopy measurements with 1.5-, 2-, and 4-μM coat-protein dimers. Data from these replicates are shown in Fig. 4A and *SI Appendix*, Fig. S5. We also perform a control experiment in which we tether only the DNA-linker strand to the coverslip surface, without RNA, and then perform an interferometric scattering microscopy measurement with 2-μM dimers. The results of this control are discussed in *SI Appendix* and shown in *SI Appendix*, Fig. S3.

Analysis of Intensity Traces. The start time for each trace is defined as the time at which the intensity reaches 0.001. To determine the growth time, we first do a linear least-squares fit to the portion of each trace that lies between the start time and the time at which the intensity

Garmann et al.

www.manaraa.com

first reaches that of a full capsid. We then estimate the time required to grow a full capsid (bind 90 dimers) by approximating the growth rate as the slope of the linear fit. Details of this analysis are described in *SI Appendix*.

TEM. We use negatively stained TEM to image the protein structures that assemble around RNA that is either tethered to a functionalized gold nanoparticle (as in Fig. 2*B* and *SI Appendix,* Fig. S4) or free in solution (as in Fig. 3*D* and *SI Appendix,* Fig. S8). The method used to functionalize the surfaces of the gold particles is similar to that used for the coverslips and is described in detail in *SI Appendix*. The assembly reaction is carried out in assembly buffer by mixing the specified micromolar concentration of coat-protein dimers with either 0.2 nM of RNA-labeled gold particles or 10 nM of free RNA. This mixture is left to sit for 10 min at room temperature and then added to a plasma-etched carbon-coated TEM gird (Ted Pella), stained with methylamine tungstate stain solution (Nanoprobes), and visualized on a Tecnai F20 (FEI) TEM operated at 120 kV. Further details on TEM sample preparation and imaging are given in *SI Appendix*.

Data Availability. All data presented are available from the corresponding author on reasonable request.

1. D. L. D. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24 (1962).
2. T. Sugiyama, R. R. Hebert, K. A. Hartman, Ribonucleoprotein complexes formed between bacteriophage MS2 RNA and MS2 protein *in vitro*. *J. Mol. Biol.* **25**, 455–463 (1967).
3. S. D. Hicks, C. L. Henley, Irreversible growth model for virus capsid assembly. *Phys. Rev. E* **74**, 031912 (2006).
4. J. M. Fox *et al.*, Comparison of the native CCMV virion with *in vitro* assembled CCMV virions by cryoelectron microscopy and image reconstruction. *Virology* **244**, 212–218 (1998).
5. P. K. Sorger, P. G. Stockley, S. C. Harrison, Structure and assembly of turnip crinkle virus: II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658 (1986).
6. S. Li, P. Roy, A. Travesset, R. Zandi, Why large icosahedral viruses need scaffolding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10971–10976 (2018).
7. R. Twarock, J. Y. Bingham, E. C. Dykeman, P. G. Stockley, A modelling paradigm for RNA virus assembly. *Curr. Opin. Virol.* **31**, 74–81 (2018).
8. J. D. Perlmutter, M. F. Hagan, The role of packaging sites in efficient and specific virus assembly. *J. Mol. Biol.* **427**, 2451–2467 (2015).
9. A. Zlotnick, Are weak protein–protein interactions the general rule in capsid assembly? *Virology* **315**, 269–274 (2003).
10. J. Carey, O. C. Uhlenbeck, Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. *Biochemistry* **22**, 2610–2615 (1983).
11. D. Beckett, H.-N. Wu, O. C. Uhlenbeck, Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**, 939–947 (1988).
12. A. Borodavka, R. Tuma, P. G. Stockley, Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15769–15774 (2012).
13. J. A. Speir, S. Munshi, G. Wang, T. S. Baker, J. E. Johnson, Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**, 63–78 (1995).
14. M. Chevreuil *et al.*, Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nat. Commun.* **9**, 3071 (2018).
15. G. Young *et al.*, Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).
16. R. F. Garmann *et al.*, A simple RNA-DNA scaffold templates the assembly of monofunctional virus-like particles. *J. Am. Chem. Soc.* **137**, 7584–7587 (2015).
17. R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, H. Reiss, Classical nucleation theory of virus capsids. *Biophys. J.* **90**, 1939–1948 (2006).
18. M. F. Hagan, "Modeling viral capsid assembly" in *Advances in Chemical Physics*, S. A. Rice, A. R. Dinner, Eds. (John Wiley & Sons, 2014), vol. 155, pp. 1–68.

19. R. F. Garmann, M. Comas-Garcia, A. Gopal, C. M. Knobler, W. M. Gelbart, The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *J. Mol. Biol.* **426**, 1050–1060 (2014).
20. W. H. Roos, R. Bruinsma, G. J. L. Wuite, Physical virology. *Nat. Phys.* **6**, 733–743 (2010).
21. R. I. Koning *et al.*, Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure *in situ*. *Nat. Commun.* **7**, 12524 (2016).
22. A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, D. Endres, A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644–14652 (1999).
23. P. E. Prevelige Jr, D. Thomas, J. King, Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824–835 (1993).
24. J. D. Perlmutter, M. R. Perkett, M. F. Hagan, Pathways for virus assembly around nucleic acids. *J. Mol. Biol.* **426**, 3148–3165 (2014).
25. G. G. Pickett, D. S. Peabody, Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein. *Nucleic Acids Res.* **21**, 4621–4626 (1993).
26. A. J. Olson, Y. H. E. Hu, E. Keinan, Chemical mimicry of viral capsid self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20731–20736 (2007).
27. H. D. Nguyen, V. S. Reddy, C. L. Brooks III, Invariant polymorphism in virus capsid assembly. *J. Am. Chem. Soc.* **131**, 2606–2614 (2009).
28. S. Li, R. Zandi, A. Travesset, G. M. Grason, Ground states of crystalline caps: Generalized jellium on curved space. *Phys. Rev. Lett.*, in press.
29. T. Hu, B. I. Shklovskii, Kinetics of viral self-assembly: Role of the single-stranded RNA antenna. *Phys. Rev. E* **75**, 051901 (2007).
30. O. Rolfsson *et al.*, Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* **428**, 431–448 (2016).
31. X. Dai *et al.*, In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116 (2017).
32. S. J. Stray *et al.*, A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8138–8143 (2005).
33. T. G. W. Edwardson, D. Hilvert, Virus-inspired function in engineered protein cages. *J. Am. Chem. Soc.* **141**, 9432–9443 (2019).
34. J. H. Strauss, R. L. Sinsheimer, Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *J. Mol. Biol.* **7**, 43–54 (1963).
35. J. Ortega-Arroyo, D. Cole, P. Kukura, Interferometric scattering microscopy and its combination with single-molecule fluorescence imaging. *Nat. Protoc.* **11**, 617–633 (2016).
36. A. M. Goldfain, R. F. Garmann, Y. Jin, Y. Lahini, V. N. Manoharan, Dynamic measurements of the position, orientation, and DNA content of individual unlabeled bacteriophages. *J. Phys. Chem. B* **120**, 6130–6138 (2016).
37. C. Joo, T. Ha, Single-molecule FRET with total internal reflection microscopy. *Cold Spring Harb. Protoc.* **2012**, 1223–1237 (2012).

www.manaraa.com