

# A Kinetic Zipper Model and the Assembly of Tobacco Mosaic Virus

Daniela J. Kraft,<sup>†\*</sup> Willem K. Kegel,<sup>†</sup> and Paul van der Schoot<sup>†§\*</sup>

<sup>†</sup>Van 't Hoff Laboratory for Physical and Colloid Chemistry, Debye Institute for NanoMaterials Science and <sup>‡</sup>Institute for Theoretical Physics, Utrecht University, Utrecht, The Netherlands; and <sup>§</sup>Department of Applied Physics, Eindhoven University of Technology, Eindhoven, The Netherlands

**ABSTRACT** We put forward a modified Zipper model inspired by the statics and dynamics of the spontaneous reconstitution of rodlike tobacco mosaic virus particles in solutions containing the coat protein and the single-stranded RNA of the virus. An important ingredient of our model is an allosteric switch associated with the binding of the first protein unit to the origin-of-assembly domain of the viral RNA. The subsequent addition and conformational switching of coat proteins to the growing capsid we believe is catalyzed by the presence of the helical arrangement of bound proteins to the RNA. The model explains why the formation of complete viruses is favored over incomplete ones, even though the process is quasi-one-dimensional in character. We numerically solve the relevant kinetic equations and show that time evolution is different for the assembly and disassembly of the virus, the former exhibiting a time lag even if all forward rate constants are equal. We find the late-stage assembly kinetics in the presence of excess protein to be governed by a single-exponential relaxation, which agrees with available experimental data on TMV reconstruction.

## INTRODUCTION

Arguably, virology as a field of study in biology started in 1898 with a publication of Beijerinck on a communicable disease afflicting tobacco plants, causing lesions in and curling of the leaves detrimental to crop yields (1–3). He showed that the disease was caused by neither a bacterium nor any other microscopic organism but by a contagious fluid that reproduces itself in diseased plants. This contagious fluid contains what we now know to be tobacco mosaic virus particles. Tobacco mosaic virus (TMV) is a rod-shaped virus consisting of a single-stranded RNA molecule of 6395 nucleotides and 2130 identical protein subunits. The latter are helically arranged around the RNA molecule, 16 1/3 proteins per turn, producing a particle that measures ~300 nm in length and 18 nm in width (4). It has a characteristic cylindrical cavity ~4-nm wide that runs through the entire length of the rod; the RNA is locked in the protein body of the cylinder ~2 nm from the cavity. Each coat protein binds to three nucleotides via a combination of hydrogen-bonding, hydrophobic, and electrostatic interactions (5).

Although superficially TMV perhaps appears like a simple virus consisting of only two components, coat protein and single-stranded RNA, the processes involved in its reconstitution from its constituents are far from that even *in vitro* (6,7). Despite half a century of intense research there are still many open questions and we refer to recent reviews of the assembly kinetics by Caspar and Namba (7), Butler (4), and Klug (6). Under conditions similar to

those where in infected tobacco cells assembly occurs, a 20S coat protein aggregate is known to be required for efficient nucleation (4,6,8). Still under debate is the precise nature of this 20S aggregate structure, that is, whether it constitutes a disk or a short helix and therefore two nucleation pathways have been put forward. See Fig. 1.

In both cases the adsorption of the first 20S protein aggregate at the origin-of-assembly sequence (OAS) of the RNA requires a conformational change of the 20S aggregate, either from a disk to a proto-helix (9,10) or from a disordered to an ordered state of the inner loops of the helical aggregate (7,11–13). The subsequent elongation of the virus is currently viewed to involve a bidirectional process originating from the OAS on, with a much more rapid elongation toward the 5' terminus than toward the 3' terminus. The latter probably occurs by the addition of single subunits (14,15), whereas the elongation toward the 5' tail may take place by the addition of disks, helices (4,9,16), and/or smaller protein aggregates (17–19). See Fig. 1. Despite the controversies surrounding the nucleation and elongation of TMV, there is little doubt that the initial conformational switching is crucial to an efficient and successful assembly of the virus. Indeed, the concept of autosteric control, in which conformational switching is used as a regulatory control mechanism in biology, and which was put forward by Caspar 30 years ago, was inspired by assembly studies on TMV (20). Strong support for the importance of conformational switching for self-assembly of viruses is not restricted to TMV and other rodlike viruses such as papaya mosaic virus (21), but extends to icosahedral viruses (22), e.g., hepatitis B virus (23), the plant virus CCMV (24), and the phage MS2 (25).

Here, we argue that some sort of allosteric conformational switching is not only advantageous to the assembly

Submitted November 1, 2011, and accepted for publication May 3, 2012.

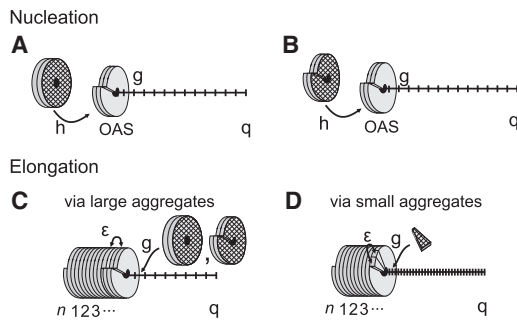
\*Correspondence: d.j.kraft@uu.nl or p.vanderschoot@phys.tue.nl

Daniela Kraft's present address is Center for Soft Matter Research, New York University, New York, NY.

Editor: Charles Wolgemuth.

© 2012 by the Biophysical Society  
0006-3495/12/06/2845/11 \$2.00

doi: 10.1016/j.bpj.2012.05.007



**FIGURE 1** Schematic representation of our kinetic zipper model for the assembly of Tobacco Mosaic Virus from coat proteins onto RNA templates with  $q$  binding sites. Nucleation of the assembly of the rodlike virus particles starts at the origin of assembly (OAS) region and is speculated to occur by either of two mechanisms: (A) by conformational switching of the 20S disk aggregate to a helix, which requires an energy  $h$ , to the OAS. (B) Alternatively, nucleation takes place by binding of the 20S helical aggregate with the RNA and ordering of the inner loops of the proteins making up the aggregate, which requires an energy  $h$ . Subsequent elongation happens by binding of (C) larger units such as the disk or the helical 20S protein aggregate or (D) the much smaller A-protein. Binding of the protein units to the RNA yields a binding free energy  $g$ , and the binding between two protein disks is associated with an interaction free energy  $\epsilon$ . This guarantees sequential assembly commencing at the OAS.

of viruses but may in fact even be a necessity if only to prevent empty viral shells from self-assembling in solution. For linear viruses, the additional reason is that the self-assembly of molecular building blocks (coat proteins) onto the linear template molecule (a single-stranded RNA molecule) is dominated by fluctuations, as this in essence constitutes a quasi-one-dimensional process (26). Fluctuations imply imperfect coverage of the template molecule, which is detrimental to the survival of the virus as the coat protects the genome from, e.g., the attack of nucleases. Indeed, recent experiments on the reversible, Langmuir-type adsorption of naphthalene derivatives to single-stranded, homopolymeric DNA molecules support this: for this kind of system, full coverage is virtually impossible to attain (27).

We base our conclusions on the application of a model inspired by the well-known zipper model originally devised to describe the melting of DNA (28). We specifically model the self-assembly of the virus from the OAS in the 5' direction and for simplicity ignore the finalization of it in the 3' direction. The latter process is slow and it involves a relatively small portion of the RNA (4). We find that allosteric regulation necessitating the sequential binding of coat proteins starting at the OAS strongly favors either complete assembly toward the 5' end or no assembly at all. The strong suppression of incompletely covered RNAs we believe is of evolutionary advantage to the virus. A kinetic version of the zipper model is proposed to describe the kinetics of reversible assembly. We find that conformational switching introduces nucleation-type kinetics, and an inherent asymmetry in the assembly and disassembly kinetics. Although highly idealized, our model describes

the limited available experimental data on the encapsulation of TMV RNAs by coat proteins reasonably well. From comparison with these data, we learn that the precise determination of the start time of assembly, as well as knowledge about the thermodynamic state at the starting and end-points of an experiment, are crucial to infer quantitative predictions from the model such as the size of the basic protein-building units.

## EQUILIBRIUM ASSEMBLY MODEL

To describe the assembly of TMV virus particles in a solution containing single-stranded RNAs and 20S aggregates, we apply a model that, in essence, is equivalent to the well-known zipper model for the melting of DNA (28). In our case it is used to describe the elongation of the virus capsid from the OAS onwards to the 5' end of the RNA. The model does not allow for random adsorption onto different sites of the template, for this is not seen in nature and in fact constitutes a cooperative binding that in itself strongly promotes complete coverage of the RNA (16,29). This is in contrast to the also well-known model of McGhee and von Hippel (26,30), in which random adsorption is allowed, and therefore also in contrast to the equivalent one-dimensional Ising model of ferromagnetism as well as the Bragg-Zimm model for the helix-coil transition (31–33).

Binding of the first protein to the template necessitates a conformational change of the proteins to accommodate a helical structure if the nucleating agent is indeed a disk (9,10), or, if it is a short helix, to induce a helix-coil transition in the so-called flexible loop involved in the locking in of the RNA (7,11–13). The helical aggregate taking part in the binding with the OAS we presume to be a high-energy structure, otherwise it would already form in free solution—that is, without binding to the RNA. This is plausible, as elongation of this structure in the absence of bound RNA is prevented, either by the closed structure of the disk, or the electrostatic repulsion between the carboxylate pairs and steric hindrance of the flexible-loop structures within the short helices (7). Of thermodynamic importance for our model is the required conformational switching and the associated free energy cost  $h(n) > 0$  that this implies for  $n = 1$ . We thus need not specify the precise structure of the nucleating agent.

We assume that the switching of subsequent protein units is catalyzed by the presence of the previous protein unit and costs much less free energy than that of the first. For simplicity, we put this free energy cost equal to zero, so we put  $h(n) = 0$  for  $n > 1$  (note that, formally, this contribution can be absorbed in the free energy gain of binding). This way, allosteric binding enters in our model in two ways: first to enforce sequential binding of protein units, and second through a nucleation barrier that will further increase the level of cooperativity of the self-assembly. Binding of the first 20S aggregate and of each subsequent

coat protein aggregate after that, shown schematically in Fig. 1, liberates free energy. This free energy is in part needed to allow for the conformational switching of the first 20S aggregate (and/or the proteins therein) upon binding. There are two obvious sources of binding free energy: that of the interaction of the proteins with the RNA, and that of the protein-protein contacts in the growing helix. The former is relative to that associated with the self-basepairing of the RNA in free solution, whereas the latter is taken relative to that between the proteins in the aggregates in free solution. Given these assumptions, we can now make the model more explicit.

Let each viral RNA strand present in the solution act as a template that can accommodate a maximum of  $q$  protein units. Considering a total of  $N$  RNA strands in the system therefore corresponds to  $Nq$  binding sites for protein material. If the total amount of protein material present in the solution corresponds to  $M$  protein units, we can define the stoichiometry between available binding sites and protein aggregates as  $\lambda \equiv Nq/M$ . It turns out useful to work not in a canonical but grand canonical ensemble, in which case  $N$  and  $M$  are expectation values. Later on we return to the canonical ensemble. The quantity of interest describing the thermodynamic state of the solution is, in that case, the grand potential. Adsorption of the protein onto the RNA liberates a binding free energy,  $g < 0$ , presumed equal for all units. This parameter also implies any helicase activity the protein might have on the RNA. Furthermore, sequential binding of protein to previously adsorbed ones produces an additional protein-protein interaction free energy,  $\epsilon < 0$ .

Let  $\Omega$  denote a dimensionless grand potential density (note that the grand potential is scaled by the thermal energy, and multiplied by the ratio of a molecular volume and the system volume) of coat protein aggregates, RNA molecules, and partially assembled viruses in an ideal solution of volume  $V$ . The only interactions we account for are those involved in the binding of protein aggregates onto the RNA binding sites. The grand canonical potential then consists of contributions from the translational entropy of protein units in solution—the RNA strands with  $n$  adsorbed protein units, the chemical potentials of RNA molecules,  $\mu_R$ , and that of the free and adsorbed protein aggregates,  $\mu_P$ . There is also a contribution accounting for the different configurations of  $n$  bound protein units on an RNA, expressed in an intrachain partition function  $Z(n)$ . If the dimensionless number densities of the coat proteins and partially assembled viruses (i.e., RNA molecules with  $n$  adsorbed protein units ( $0 \leq n \leq q$ )) are defined as  $\rho_P$  and  $\rho_R(n)$ , we have

$$\Omega = \sum_{n=0}^q \rho(n)_R [\ln \rho_R(n) - 1 - \ln Z(n) - \mu_R - n\mu_P] + \rho_P (\ln \rho_P - 1 - \mu_P). \quad (1)$$

In thermodynamic equilibrium, the optimal number densities of the RNA molecules and the free protein units minimize the grand potential. Hence, we set

$$\left( \frac{\partial \Omega}{\partial \rho_R(n)} \right)_{\mu_R, \mu_P, V, T} = 0, \quad (2)$$

and find that the number density of chains that have  $n$  bound protein aggregates is given by

$$\rho_R(n) = Z(n) \exp(\mu_R + n\mu_P) \quad (3)$$

and the number density of free protein units by  $\rho_P = \exp(\mu_P)$ . Given the above assumptions, the dimensionless free energy  $F(n)$  of  $n$  protein aggregates adsorbed to one RNA can be written as

$$\begin{aligned} F(n) &= h + \epsilon(n-1) + gn, & \text{for } n \geq 1 \\ F(0) &= 0, & \text{for } n = 0. \end{aligned} \quad (4)$$

Here,  $h(n=1) > 0$  is the conformational free energy cost of the first 20S aggregate upon adsorption onto the OAS, and for the subsequently adsorbed protein units we set  $h(n > 1) = 0$ . The expression  $\epsilon < 0$  is the protein-protein interaction free energy and  $g < 0$  is the free energy associated with the binding of the protein units to the RNA. We take all energies to be in units of the thermal energy,  $k_B T$ .

It turns out useful not to focus on the canonical partition function  $Z(n)$ , but rather on the semigrand canonical partition function

$$\Xi(\mu_P) = \sum_{n=0}^q \exp[-F(n) + n\mu_P], \quad (5)$$

that is, the sum over all possible configurations of bound coat protein aggregates onto RNA. Here,  $F(n) = -\ln Z(n)$ . The quantity  $\Xi$  is easily calculated using Eq. 4 and reads

$$\Xi = 1 + \sigma s \left[ \frac{1 - s^q}{1 - s} \right], \quad (6)$$

where we have defined the parameters  $s \equiv \exp(-\epsilon - g + \mu_P) = \rho_P \exp(-\epsilon - g)$  and  $\sigma \equiv \exp(-h + \epsilon)$ . The quantity  $s$  is a measure for the affinity of proteins for RNA molecules, combining the binding free energy between the proteins and that to the RNA with the number density of free protein aggregates. The effect of nucleation and allostery is captured in the quantity  $\sigma$ , a measure for the net free energy cost of adsorption of the first 20S aggregate as described in the introduction. If  $s \cdot \sigma$  is small, then the binding of the first 20S aggregate to the template constitutes a high-energy state.

From the semigrand partition function of the RNA molecules, we can now calculate the average level of their coverage by the coat proteins,

$$\langle \theta \rangle = \sum_{n=0}^q \frac{n \rho_R(n)}{q} \sum_{n=0}^q \rho_R(n) = q^{-1} \left( \frac{\partial \ln \Xi}{\partial \mu_P} \right)_{\mu_R, V, T}, \quad (7)$$

and find

$$\langle \theta \rangle = \frac{\sigma}{q} \times \frac{s}{1-s} \times \frac{1 - (q+1)s^q + qs^{q+1}}{1-s + \sigma s(1-s^q)}. \quad (8)$$

Equation 8 shows that increasing the affinity of the protein for the RNA, that is, increasing the value of  $s$ , e.g., by changing the pH or the concentration of free protein aggregates in solution (5), can only lead to full encapsidation of the viral RNA and  $\langle \theta \rangle \rightarrow 1$  if  $s \gg 1$ . For  $s \ll 1$ ,  $\langle \theta \rangle \rightarrow 0$  and RNA molecules and protein aggregates then remain freely dispersed in solution. For  $s = 1$ , we have  $\langle \theta \rangle = 1/2 \sigma(q+1)/(1+q)$ , which is equal to one-half if  $q\sigma \gg 1$  or to  $1/2 \sigma(q+1)$  if  $q\sigma \ll 1$ .

From the plots of  $\langle \theta \rangle$  in Fig. 2, C and D, it is apparent that the transition from free RNA and protein aggregates to fully functional TMV particles occurs for values close to  $s = 1$  and becomes sharper with increasing template length. The required sequential binding implies long-range correlations between bound protein units, which is why the adsorption transition becomes a true phase transition in the limit  $q \rightarrow \infty$  despite the model being one-dimensional. For templated assembly models such as that based on the one-dimensional Ising model (26), there is no phase transition even in the infinite-chain limit. Allosterism causes the point where half of the RNA molecules are, on average, covered with coat protein (the adsorption transition point) to shift toward larger values of  $s$ , while at the same time sharpening the transition (see Fig. 2, C and D).

We now investigate whether the mean coverage applies to all individual RNA molecules or whether there are great differences in coverage between different ones. From the equilibrium distribution of RNAs occupied by  $n$  protein aggregates, defined as

$$P(n) = \frac{\rho_R(n)}{\sum_{n=0}^q \rho_R(n)}, \quad (9)$$

we find

$$P_{eq}(0) = \frac{1}{\Xi} \quad \text{for } n = 0$$

$$P_{eq}(n) = \frac{\sigma^n s^n}{\Xi} \quad \text{for } n > 0. \quad (10)$$

From Fig. 2, A and B, we conclude that allosterism plays an important role in strongly favoring fully, over partially, covered RNA molecules. For  $\sigma = 1$ , that is, in the absence of allosteric effects, RNA molecules coated with  $n$  protein units obey an exponential distribution in  $n$ , leading to

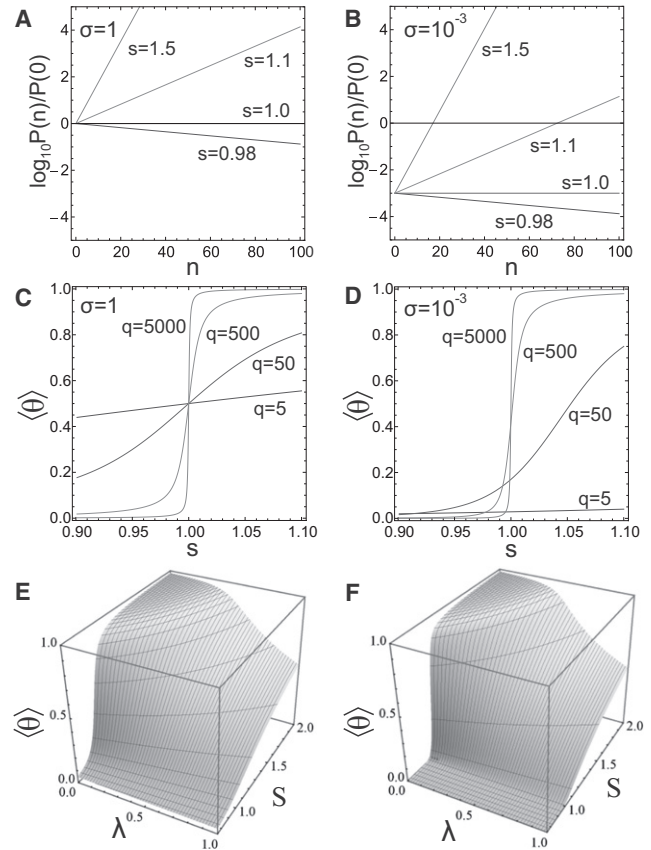


FIGURE 2 Ratio of probabilities  $P(n)/P(0)$  for an RNA molecule to be coated by  $n$  protein units as a function of the affinity  $s$  (A) in the absence of allosterism ( $\sigma = 1$ ), and (B) in the presence of allosteric effects ( $\sigma = 10^{-3}$ ). The average fraction of binding sites occupied by protein aggregates,  $\langle \theta \rangle$ , for (C)  $\sigma = 1$  and (D)  $\sigma = 10^{-3}$  also as a function of the affinity  $s$ . For the number of binding sites on each polymeric template we set  $q = 63$ . (E and F) The average level of coverage of RNA strands by coat proteins  $\langle \theta \rangle$  as a function of the bare affinity  $s$  (E) in the absence of allosterism ( $\sigma = 1$ ), and (F) in the presence of allosteric effects ( $\sigma = 10^{-3}$ ). The effective number of binding sites equals  $q = 63$ .

partially covered TMV particles of all intermediate lengths  $n$  for  $s > 1$  albeit favoring the completely filled one for  $n = q$ . When allosterism does come into play and  $\sigma \ll 1$ , the probability for coverage of RNA with coat proteins below a critical value  $0 < n^* = -\ln \sigma / \ln s$  is much lower than that of the empty RNAs (with  $n = 0$ ). Only above  $n^*$ , substantial coverage of RNA with coat protein occurs, producing either empty or almost fully covered viruses, in particular if  $\sigma$  is sufficiently small. Nucleated virus capsids with  $n < n^*$  are thus high-energy structures that do not form free in solution in appreciable concentrations and are thermodynamically unfavorable.

We now come back to the issue of linking the affinity,  $s$ , to an experimentally controllable quantity and return from a grand canonical to a canonical description in which the amount of material in solution is fixed rather than the chemical potentials. Let  $\phi_P$  denote the overall (dimensionless) concentration of proteins present in the solution. From the

condition of the conservation of mass of the protein aggregates, we have

$$\phi_P = \rho_P + \sum_{n=0}^q n \rho_R(n) = \rho_P + q \langle \theta \rangle \rho_R, \quad (11)$$

where  $\rho_R = \sum_{n=0}^q \rho(n)$  is the dimensionless RNA concentration. Hence,

$$S = s + S \lambda \langle \theta \rangle, \quad (12)$$

with  $\lambda = q \rho_R / \phi_P$  the earlier defined stoichiometric ratio and  $S \equiv \phi_P \exp(-\epsilon - g)$  a bare affinity. The bare affinity is the product of the overall protein concentration  $\phi_P$ , which is a known quantity, and a binding constant  $K \equiv \exp(-\epsilon - g)$ . Clearly, Eq. 12 is an implicit expression that needs to be solved for  $s$ . For the case of excess protein  $\lambda \rightarrow 0$  we obviously have the identity  $s = S$ . In the other limit, for concentrations of RNA binding sites far in excess of that of protein units in the solution,  $\lambda \rightarrow 0$  and  $s \rightarrow 0$ , implying low average RNA coverage.

Perhaps not entirely surprisingly, we conclude that to fully cover all the RNA molecules present in the solution such that  $\langle \theta \rangle = 1$ , we need an amount of protein far in excess of the number of binding sites,  $qN$ . This can also be deduced from Fig. 2, E and F, where we plot the mean fraction of occupied sites as a function of the bare affinity,  $S$ , and of the stoichiometry,  $\lambda$ . Apparently, the level of cooperativity of the binding of the proteins to the RNA strongly decreases with increasing stoichiometry. It is in this context not quite clear why so many experimental studies focus on stoichiometric ratios close to unity (13,17,34,35).

## KINETIC MODEL

To extend our equilibrium model toward a description of the dynamics of the assembly and disassembly of the rodlike virus particles, we consider a reversible sequential association or dissociation of coat proteins, with adsorption rates  $k_+(n)$  and desorption rates  $k_-(n)$  that depend on how many protein units  $n$  are bound onto the RNA. From this, we obtain a set of kinetic equations for the probabilities  $P(n, t)$  that RNA molecules are occupied by  $n$  protein aggregates at time  $t$ . Our approach is similar in spirit, yet not directly comparable, to that of recent work describing the assembly and disassembly kinetics of capsids of spherical viruses (36–38), and approaches to describe the helix-coil transition kinetics in polypeptides (31–33).

For convenience, we imply the time dependence of the occupation probabilities by writing  $P(n) = P(n, t)$ , and distinguish these from the equilibrium distribution written as  $P_{eq}(n)$ . The set of kinetic equations is

$$\frac{\partial P(0)}{\partial t} = -k_+(0)P(0) + k_-(1)P(1), \quad (13)$$

$$\begin{aligned} \frac{\partial P(m)}{\partial t} = & -k_+(m)P(m) + k_-(m+1)P(m+1) - k_-(m)P(m) \\ & + k_+(m-1)P(m-1) \text{ for } 0 < m < q, \end{aligned} \quad (14)$$

$$\frac{\partial P(q)}{\partial t} = -k_-(q)P(q) + k_+(q-1)P(q-1), \quad (15)$$

where the first and last equation involve the empty and fully bound RNAs, respectively. In time, the system evolves toward its equilibrium state as discussed in the previous section, and hence  $\lim_{t \rightarrow \infty} P(n, t) = P_{eq}(n)$  and  $\lim_{t \rightarrow \infty} \partial P(n, t) / \partial t = 0$ . With these conditions, the disassembly rates  $k_-(n)$  can be expressed by the assembly rates  $k_+(n)$  and the equilibrium probabilities  $P_{eq}(n)$  as  $k_-(n) = k_+(n-1)P_{eq}(n-1)/P_{eq}(n)$ , where the equilibrium distribution  $P_{eq}(n)$  is given by Eq. 10.

After the adsorption of the first 20S aggregate onto the OAS with a rate  $k_+(0)$ , the adsorption kinetics of subsequent protein aggregates should not differ significantly from each other. We therefore presume that the on-rates are independent of the number of previously adsorbed protein units, so  $k_+(n) = k_+$ , at least for  $n > 1$ . Then, it makes sense to relate the on-rate of the first protein aggregate to that of the others,  $k_+(0) \equiv \kappa k_+$ . Here,  $\kappa$  is a measure for how much faster or slower binding of the first 20S aggregate is relative to that of the subsequent protein units. Clearly, if  $\kappa \ll 1$ , we expect assembly to be slowed down drastically as the binding of the first aggregate will become rate-limiting. However, as we shall see below, for the viruses to assemble, we need to pass a free energy barrier if  $\sigma \ll 1$ . The reason is, of course, that assemblies of size  $0 < n < n^*$  are high-energy structures, as discussed in the preceding section. This implies that in practice it will be difficult to distinguish between kinetic and thermodynamic nucleation, although their effect on the disassembly kinetics turns out to be quite different.

If, in addition, we substitute  $k_+ t \equiv \tau$  and  $0 \leq f(n) \equiv P(n)/P_{eq}(n) \leq 1$ , we obtain the kinetic equations

$$\frac{\partial f(0)}{\partial \tau} = -\kappa [f(0) - f(1)], \quad (16)$$

$$\frac{\partial f(1)}{\partial \tau} = -f(1) + f(2) + \frac{\kappa}{\sigma S} [f(0) - f(1)], \quad (17)$$

$$\begin{aligned} \frac{\partial f(m)}{\partial \tau} = & -f(m) + f(m+1) + \frac{1}{S} [f(m-1) \\ & - f(m)], \text{ for } 1 < m < q, \end{aligned} \quad (18)$$

$$\frac{\partial f(q)}{\partial \tau} = \frac{1}{S} [f(q-1) - f(q)], \quad (19)$$

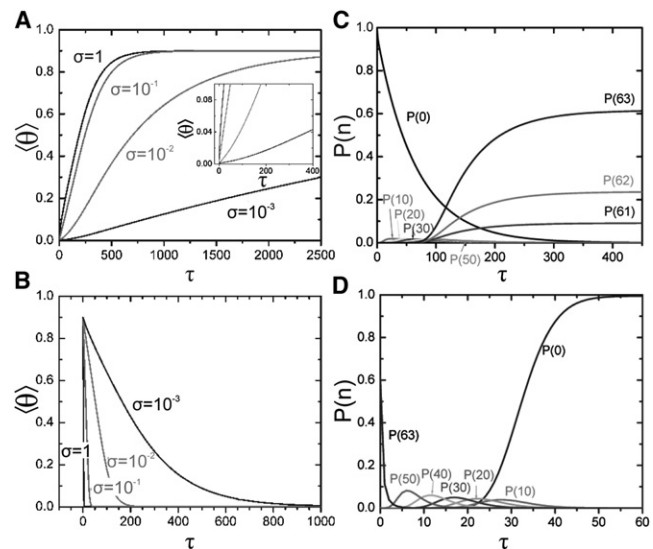
where we see that the second equation depends on the ratio of  $\kappa$  and  $\sigma$ . The former is a measure for the importance of kinetic nucleation and the latter for thermodynamic nucleation. Note that  $\kappa$  occurs in the first equation independently of  $\sigma$ , yet the ratio of  $\kappa$  and  $\sigma$  occurs in the kinetic equation of RNA molecules with one bound protein aggregate. For the matrix notation, see the [Supporting Material](#).

We employed a classical fourth-order Runge-Kutta method to numerically solve the kinetic equations for protein concentrations in excess of the number of binding sites, i.e.,  $\lambda \rightarrow 0$  or  $S = s$ , and investigate the characteristics of assembly and disassembly of our model. We need not focus on the case of excess protein, of course. Subunit depletion can be taken into account by calculating  $\langle \theta \rangle$  at every time step  $\tau$  and calculating from [Eq. 11](#) the corresponding value of  $s$  that becomes a function of  $\tau$ . The physics is not appreciably different, which is why we focus our discussion on the numerically much simpler case  $\lambda \rightarrow 0$ .

Assembly from nearly empty to fully covered RNA molecules requires a deep quench in the affinity  $S$ , for example by a sudden change in pH or temperature, or by a sudden addition of proteins to a solution of RNA molecules. Upon a quench from, say, conditions where  $\langle \theta \rangle = 1.0 \times 10^{-4}$  to those with  $\langle \theta \rangle = 0.99$ , assembly occurs via unstable intermediates toward the stable equilibrium distribution. See [Fig. 3, C and D](#). In this case, the majority of RNA molecules, roughly 95%, is covered with  $>60$  protein 20S assemblies after assembly has completed, if we presume  $q = 63$  (note that  $q = 63$  corresponds to full coverage of the RNA if we presume the adsorbing protein units to consist of 20S assemblies consisting of  $\sim 17$  coat proteins; we chose this value of  $q$  only for discussing the model and do not imply that the building blocks have to be 20S units) and  $\sigma = 0.01$ . For these values of  $\sigma$  and  $q$ , the affinities  $s = 0.35$  and  $s = 2.6$  correspond to the values for the initial and final coverage.

The rate-determining step for assembly is the adsorption and conformational switching of the first 20S protein aggregate, as indicated by the slow decrease in concentration of empty RNA molecules,  $P(0)$ . See [Fig. 3 C](#). After nucleation has occurred and the highly unfavorable intermediate states are populated temporarily, assembly occurs quickly toward fully assembled virus particles. If allosteric effects were not important and  $\sigma = 1$ ,  $P(0)$  decreases significantly more quickly than if  $\sigma \ll 1$ . So, although allostery helps to completely cover the RNAs, it does so at the cost of slowing down the assembly kinetics.

Not surprisingly, the disassembly is not delayed by a sluggish nucleation process and therefore the fraction of fully assembled viruses quickly decreases upon a downward quench of the affinity  $s$ . According to our model, disassembly is quite a bit swifter than assembly. Still,  $\sigma$  influences the disassembly kinetics, although subtly (see [Fig. 3 B](#)). Unfortunately, we have not been able to find in vitro disassembly studies in the literature, so whether this prediction holds any water remains unclear. It is true that in the in vivo exper-



**FIGURE 3** (A) Assembly and (B) disassembly kinetics of TMV as modeled by a kinetic zipper for different values of the allosteric parameter  $\sigma$  presuming 63 binding sites for each RNA molecule. The quenches are between average surface coverages  $\langle \theta \rangle = 0.001 \rightarrow \langle \theta \rangle = 0.9$  and  $\langle \theta \rangle = 0.9 \rightarrow \langle \theta \rangle = 0.001$ . (A, *Inset*) Expanded view of the onset of assembly, to point out the lag time introduced by the allosteric factor  $\sigma$ . (C) Assembly after a sudden quench from affinities  $s_1 = 0.35 \rightarrow s_2 = 2.6$  and (D) disassembly after the quench  $s_2 = 2.6 \rightarrow s_1 = 0.35$ . Shown are the probabilities  $P(n)$  of  $n$  protein units on RNA molecules consisting of 63 binding sites as a function of the dimensionless time  $\tau$ . Here,  $s_1 = 0.35$  and  $s_2 = 2.6$  correspond to a coverage of  $\langle \theta \rangle = 10^{-4}$  and  $\langle \theta \rangle = 99$ , respectively, and the allosteric parameter  $\sigma = 0.01$ . All on-rates are equal, implying  $\kappa = 1$ .

iments, disassembly from the 5' to the 3' end occurs rapidly in 2–3 min (39–41). In vitro assembly into full viruses occurs on the timescale of 6–10 min, which would support our prediction (4). However, disassembly in vivo is thought not to proceed via spontaneous processes but rather by cotranslational disassembly (42,43) so this observation does not lend actual support for our model.

Decreasing the value of the allostery parameter  $\sigma$  increases the influence of allosteric effects on the binding of protein to the RNA and thus the dynamics. As can be seen from [Fig. 3 A](#), the average coverage by coat proteins shows a lag time in the early-stage kinetics due to the nucleation step that strongly influences the overall rate of assembly. The rate of disassembly is decreased as well by a stronger allosteric effect, as is shown in [Fig. 3 B](#). Here, assembly and disassembly were chosen between a coverage of  $\langle \theta \rangle = 0.001$  and of  $\langle \theta \rangle = 0.9$ . It implies that different quench depths  $\Delta s$  were employed because the coverage depends on  $\sigma$ , as discussed in the previous section.

The disassembly rates for deep quenches of equal quench depth are essentially equivalent for different values of  $\sigma$ , except for the very late stages (see [Fig. S1](#) in the [Supporting Material](#)). For shallow quenches in the affinity  $s$ , the rates are indeed different and the late-stage assembly can be described by a single exponential relaxation (see [Fig. S2](#)). We quantify the rates by evaluating the half-coverage time

$\tau_{1/2}$ , i.e., the time required to achieve 50% coverage (see the [Supporting Material](#) for extended discussion). We find for shallow quenches that the nucleation step predominates over the overall rate of assembly, and less so for disassembly. As a consequence, disassembly rates are always larger than assembly rates. This qualitatively different behavior for shallow and deep quenches originates from the variation of mean coverage  $\langle \theta \rangle$  with the affinity  $s$ . For deep quenches in  $s$ ,  $\langle \theta \rangle$  does not significantly change with different  $\sigma$ . For shallow quenches,  $\langle \theta \rangle$  varies strongly as a function of  $\sigma$ , which translates itself in different rates of assembly and disassembly.

In [Fig. 3](#), the assembly rates for adsorption of the first and subsequent protein units were taken to be equal, i.e.,  $\kappa = 1$  so  $k_+(0) = k_+$ . The observed nucleation-type assembly kinetics originates solely from the costly conformational switching of the first adsorbed protein aggregate and the melting of the RNA strand. By taking a different assembly rate for the adsorption of the first protein unit, implying  $\kappa \neq 1$ , or, in other words,  $k_+(0) \neq k_+$ , we find an even stronger delay of the assembly the smaller  $\kappa$  is, and the slower the first step is in comparison with subsequent steps. See [Fig. S3 A](#). The disassembly kinetics of the removal of all but the last coat protein unit is unaffected by the parameter choice of  $\kappa \ll 1$  (see the [Supporting Material](#) for details, and [Fig. S3 B](#)). Setting  $\kappa > 1$  does not influence the assembly and disassembly kinetics to any discernible level. This, we believe is caused by the presence of the high-energy intermediates, and hence to be an effect of the allostery. As the assembly kinetics is already well captured by the introduction of the allosteric factor  $\sigma$ , it seems sensible to set  $\kappa = 1$  for comparison with experiments in the following section.

## COMPARISON WITH EXPERIMENTAL DATA

From comparison with experiments we ideally would not only confirm the validity of our model but also determine the correct model parameters and thus be able to distinguish, for example, between the contested building blocks involved in TMV assembly, that is, the template length  $q$ , as well as the cooperativity  $\sigma$ . Although the model describes the dynamics of the assembly quite accurately as we shall see, the determined parameters are of limited precision due to several experimental challenges, not least the fact that RNA extracted from TMV-infected plants often contains shorter RNA pieces, presumably due to the presence of nucleases in solution. As a result, the observed length distribution of assembled virus particles contains a significant fraction of short virus particles, which are indistinguishable from incompletely assembled viruses. Furthermore, because the allostery parameter  $\sigma$  in particular influences the initial stages of the virus particle growth, determination of the starting time of the experiment is crucial yet not always clearly stated in experiments. Lastly, knowledge of the phase diagram of the virus assembly and

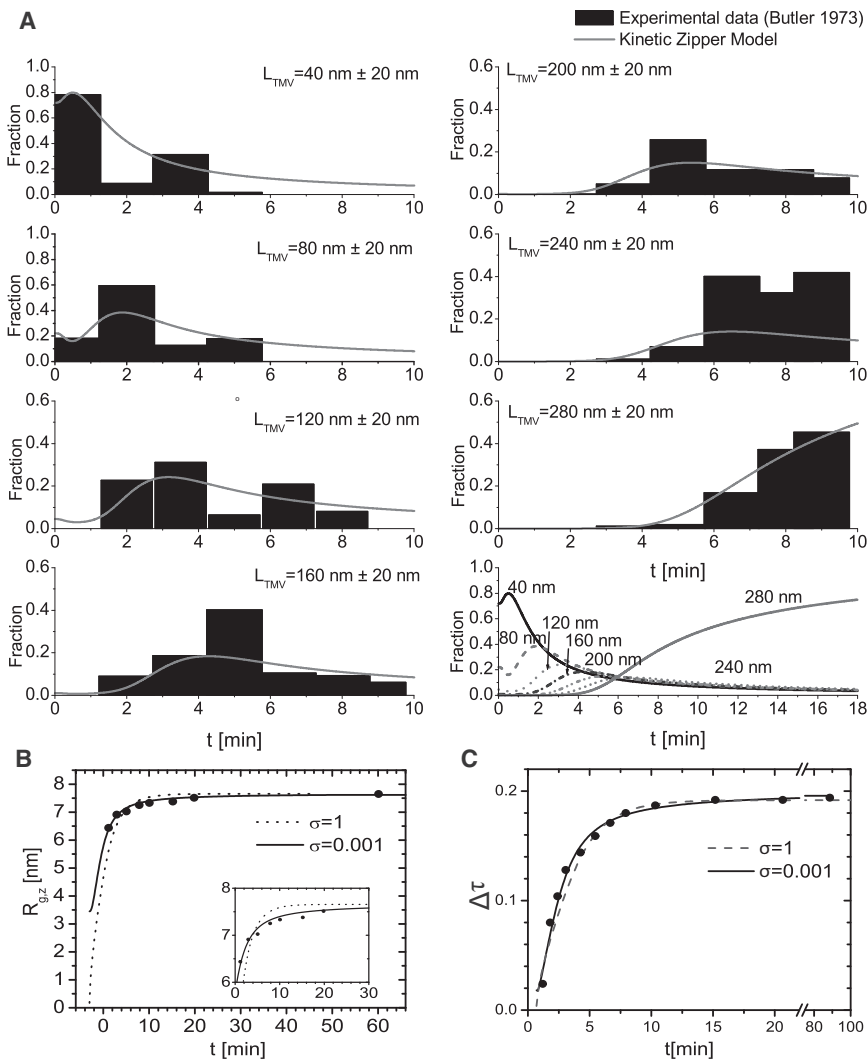
thus the experimentally employed quench depth would limit the number of freely adjustable parameters significantly. Without knowledge of the quench details, the parameter space is considerable and requires one to guess for at least some of the parameters.

To compare our model with available experimental data, we presume that the addition rate of the first protein unit is equal to that for all subsequent protein aggregates and set  $\kappa = 1$ . Even if this is not quite true and, say,  $\kappa < 1$ , the nucleation-type of kinetics that this produces cannot easily be distinguished from that resulting from the effects of allostery as we have seen previously. We thus absorb potential kinetic effects into the thermodynamics of allostery, i.e., into the value of  $\sigma$  that becomes an effective one.

The most detailed experimental data we confront our theory with are those of Butler and Finch (44), who, in 1973, studied the reconstitution of TMV particles. In these experiments, solutions of protein disks were added to the RNA molecules in excess of the number of binding sites on the RNA with a stoichiometry of  $\lambda = 0.45$ . The length distribution of the assembled structures was recorded as a function of time and corrected for the effects of the degradation of RNA molecules. Partially assembled rods shorter than 20 nm were ignored due to limitations of the electron microscopy that was used to determine the size distribution. We set the length of the template equal to the effective number of binding sites per protein unit on the viral RNA. We now presume that such a protein unit is equivalent to a 20S aggregate, i.e., either a disk or a proto-helix, as the authors have claimed to have used 20S protein aggregates. Thus, we set  $q = 63$ .

The best fit to the data, presented in [Fig. 4 A](#), we found by setting  $\sigma = 0.001$ , and presuming a quench from an affinity of  $s = 0.84$  to that of  $s = 2.6$ . This corresponds to bare values of the affinities of  $S = 0.84$  and  $S = 4.7$  and average coverage  $\langle \theta \rangle = 5 \times 10^{-4}$  and  $\langle \theta \rangle = 0.99$ . Following the experiments, we left out partially assembled rods shorter than 20 nm, equivalent to  $0 \leq n \leq 4$ , from the analysis, and we renormalized the probabilities. The timescale  $\tau = 15t$  [min] we set equal to the actual time in minutes  $t$  [min] to obtain the best fit, giving an add-on rate of  $k_+ = 15$  [min $^{-1}$ ]. Here, we presumed that the experimental data at  $t = 0$  [min] indeed corresponds to the beginning of the assembly process, that is,  $t_0 = 0$  [min] ignoring any delay time between the moment the sample was taken from the protein RNA solution, and the time at which the assembly was actually stopped. A somewhat different choice of zero-time may lead to a smaller value for  $\sigma$  than we estimated from our curve fitting.

Butler and Finch (44) divided the experimental data for the time evolution of the various levels of completion of the encapsulation of the RNA into bins of multiples of 40-nm particle lengths. As is clear from [Fig. 4 A](#), our kinetic zipper model describes the time dependence of the entire population of TMV particles of different states of



**FIGURE 4** Comparison of the predictions of the kinetic zipper model with experimental data. (A) Experiments of Butler and Finch (44) on the reconstitution kinetics of TMV. Shown are fractions in percentile units of binned length ranges as a function of dimensionless time  $\tau$ . To obtain a good fit, we set the model parameters described in the main text at  $\sigma = 0.001$  and  $\kappa = 1$ , and presumed a quench from  $s_1 = 0.84$  to  $s_2 = 2.58$ . (B and C) Comparison between experimental results probing some average of the size of the growing virus particles in TMV assembly experiments and the kinetic zipper model. (B) Radius of gyration  $R_{g,z}$ , as obtained by SAXS versus time. (Symbols) Experimental data (35). (Drawn line) Theory ( $k_+(\sigma = 1) = 40[\text{min}^{-1}]$ ,  $k_+(\sigma = 0.001) = 65[\text{min}^{-1}]$ ). (Inset) Enlarged part of the same plot. (C) Assembly kinetics followed by the increase in turbidity upon mixing TMV protein disks with TMV RNA. (Symbols) Experimental data (45). (Drawn line) Theory ( $k_+(\sigma = 1) = 60[\text{min}^{-1}]$ ,  $k_+(\sigma = 0.001) = 112[\text{min}^{-1}]$ ). Model parameters for panels B and C were set to  $\sigma = 0.001$ ,  $\kappa = 1$ . A quench from  $\langle \theta \rangle = 0$  to  $\langle \theta \rangle = 0.9$  was presumed.

completion reasonably well. In the late stages, i.e., after 10 min, still a very large fraction of almost complete TMV particles with an average length of  $240 \pm 20$  nm remains, which does not quite agree with our predictions as is evident from Fig. 4 A.

Although Butler and Finch did correct their data for growth termination due to degraded RNA, it is reasonable to assume that, in the late assembly stages, distinguishing discontinued growth of the  $240 \pm 20$  nm particles for thermodynamic reasons from that due to broken RNA molecules is not possible. The latter is of course not accounted for in our model. From the best fit to the experimental data, we conclude that the assembly has not completed at the point when the experiments halted (Fig. 4 A, bottom right).

Another explanation for the discrepancy may be that the elongation of the virus is bidirectional: assembly from the origin of assembly region toward the 5' end of the viral RNA completes much more rapidly than the assembly in the direction of the 3' end. If the assembly in the 5' direction is completed, the virus particles are roughly 260 nm in

length, which is in good agreement with the large fraction of almost fully complete viruses  $240 \pm 20$  nm in length seen in experiments. Similar experiments by Fukuda et al. (14) with a higher resolution in the particle length observed this slower assembly toward the 3' end. Although a different TMV strain (namely, the Japanese common strain OM) and higher salt concentrations were used, the length distributions during the reconstitution reaction closely resemble the ones from Butler and Finch.

We compare the evolution of the particle distribution obtained from our model for various values of  $\sigma$  and quench depths in Fig. S4 and show the best fit we could find for a template length  $q = 400$  corresponding to a smaller protein unit (A-protein, with 4–6 proteins per unit) in Fig. S5. From comparing the best fits for  $q = 63$  in Fig. 4 A and  $q = 400$  in Fig. S5, we find that our model is not capable of distinguishing between the two values for the template length and hence the building block size given the noise in the experimental data. To be able to infer the type of building blocks from similar experiments, one would



need a higher resolution of the virus particle-length distribution as well as a reasonably accurate zero-time to obtain a correct value for  $\sigma$ .

Despite all this, it seems that our model captures the main elements of the assembly kinetics of TMV, certainly of that in the direction of the 5' end. As a further test, we compared our model with reconstitution data obtained by means of small-angle x-ray scattering (SAXS) (35) and turbidity measurements (45). In both these studies, experiments were conducted with just as many coat proteins as RNA binding sites available, implying that  $\lambda = 1$ . Strictly speaking, our numerical model does not apply for this stoichiometric ratio as we presumed  $\lambda \ll 1$ , except in the initial stages when the solution concentration of protein units remains close to its initial value, and thus our predictions for the late stages may be more tenuous.

The mean radius of gyration,  $R_{g,z}$ , of the assembling virus particles, extracted from (time-resolved) static-radiation-scattering experiments is a so-called  $z$  average of all particles in the population and related to the  $z$ -averaged particle length,  $\langle L_z \rangle$ , via

$$R_g^2 = \frac{\langle L^2 \rangle_z}{12} \equiv \frac{\sum_{n=1}^q P(n)L(n)^4}{12 \sum_{n=1}^q P(n)L(n)^2}, \quad (20)$$

where  $L(n) = 1 \times n$  obviously depends on how many protein units  $n$  of length  $l$  have adsorbed onto the RNA (46) and  $\langle \dots \rangle_z$  denote  $z$  averages. In Fig. 4 B, we fit the theory to the SAXS data of Sano et al. (35), where we set  $q = 63$ ,  $l = 4.76$  [nm],  $\sigma = 0.001$ , and  $k_+ = 65$  [ $\text{min}^{-1}$ ] and presume a quench from  $\langle \theta \rangle = 0$  to  $\langle \theta \rangle = 0.9$ . The value of the add-on rate  $k_+$  that we now find is larger by a factor 4 than the one we found previously, but this may of course be due to different solution conditions. For comparison we plot the numerical results for  $\sigma = 1$  and  $k_+ = 40$  [ $\text{min}^{-1}$ ]. As the zero-time is not precisely known in these experiments either, we set the assembly to commence at  $t_0 = -3$  [min] in our model to obtain good agreement with the experimental data.

Quite similar results for the size of the growing rods may be obtained from turbidity measurements. The turbidity is proportional to the scattered total intensity that within the Rayleigh-Debye-Gans approximation scales as the weight average of the rod length (47),

$$\langle L \rangle_w = \frac{\sum_{n=1}^q P(n)L(n)^2}{\sum_{n=1}^q P(n)L(n)}. \quad (21)$$

By setting  $\sigma = 0.001$ ,  $k_+ = 112$  [ $\text{min}^{-1}$ ],  $t_0 = 0.7$  min, and presuming the assembly proceeds from a coverage of  $\langle \theta \rangle = 0$  to that of  $\langle \theta \rangle = 0.9$ , we find good agreement

between the experimental data, as is shown in Fig. 4 C. We also compare the data to the best fit for  $\sigma = 1.0$ ,  $k = 60$  [ $\text{min}^{-1}$ ],  $t_0 = 0.7$  min. For short timescales, the theoretical values for  $\sigma = 1$  diverge slightly from the experimental data.

Because the parameters associated with the best fit are somewhat ambiguous due to the experimental uncertainties, the question as to which building blocks may be involved in the assembly of TMV remains unanswered at this point. New experiments could improve the accuracy of the measurements and thus of the extracted parameters, and might allow us to distinguish between the contested assembly paths.

## CONCLUSIONS

In summary, we have presented a kinetic zipper model to describe the statics and dynamics of the assembly and disassembly of tobacco mosaic virus particles in solutions containing coat protein units and RNA molecules. The key ingredient of our theory is the integration of allostery: only the binding of the first 20S aggregate and concomitant conformational switching is penalized by a free energy cost. The conformational switching of subsequent aggregates (or single proteins) bound to the RNA is in our model catalyzed by the already present helical arrangement of proteins.

The model, although admittedly crude, does explain the main features of the *in vitro* reconstitution of TMV particles as observed in experiments. We find that allostery, if sufficiently strong, leads to all-or-nothing behavior in the assembly. This means that rather than incompletely encapsulating all RNAs in the solution, there is a strong driving force to (nearly) completely cover a fraction of RNAs, while keeping the remainder naked. Arguably, this all-or-nothing coverage enhances the survival probability of the virus, because the exposed (i.e., naked) RNA of an incomplete virus is susceptible to attack, e.g., by nucleases.

We find that the cooperative assembly of fully covered RNA molecules, i.e., intact virus particles, is strongly influenced by the stoichiometric ratio that is proportional to the number of nucleotides and proteins in the solution. Only for the case where protein is available vastly in excess of the total number of nucleotides, full coverage can be achieved.

If allostery is indeed an important mechanism in virus assembly, as we believe is the case, then our calculations show that the assembly kinetics must occur via high-energy intermediate states that are present only in low concentrations. This makes the nucleation process the rate-limiting step. The associated lag time before assembly commences seems to be influenced by both the relative rate of adsorption of the first 20S aggregate and the extent of allostery. For disassembly of the virus, these parameters only influence desorption of the last protein unit and hence not the overall rate of disassembly. The combination of conformational switching and template-assisted self-assembly is not

specific to the in vitro assembly of tobacco mosaic virus, and our model hence may apply to other viruses and protein-macromolecule binding processes as well. The late-stage assembly and the early-stage disassembly kinetics we find to be governed by a single-exponential relaxation, at least for the situation of excess protein.

Comparison of our quite simple model with available experimental data shows remarkably good agreement. From our curve fitting, we obtain estimates for the allosteric parameter of  $\sigma \approx 0.001$ . However, this value may be somewhat ambiguous related to the difficulty of establishing zero time and/or a stoichiometry ratio that is not sufficiently small as well as to experimental difficulties involving partially broken RNA pieces. Thus, to draw reliable conclusions new experiments are required, ideally conducted with the now available high-resolution electron microscopy techniques at well-controlled conditions, e.g., pH and salt concentration, and possibly varying stoichiometry. Good control over the starting time  $t_0$  is crucial for extracting an accurate value for  $\sigma$ .

An ideal testing ground for the model would be experiments employing poly-A nucleic acid containing the origin of assembly region of TMV. These experiments could distinguish between ideal assembly kinetics and any issues related to the tertiary structure of the RNA or the bidirectional assembly of TMV. Variation of the length of the poly-A molecule allows us to investigate the influence of the template length and if conducted under controlled conditions may resolve the controversy on the primary building block for assembly. Clearly, more experiments are needed to resolve these issues, as is an analysis of the kinetic model that deals with more realistic stoichiometries. The latter is left for future work.

## SUPPORTING MATERIAL

List of symbols, five figures, the kinetic equations in matrix form, additional sections, and reference (48) are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(12\)00560-7](http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00560-7).

## REFERENCES

1. Beijerinck, M. 1898. About a *contagium vivum fluidum* as the cause of the mottling of tobacco leaves [Over een *contagium vivum fluidum* als oorzaak van de vlekziekte der tabaksbladen]. *Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen*. 65:3–21.
2. Scholthof, K., J. Shaw, ..., L. Sindelar. 1999. Tobacco Mosaic Virus: One Hundred Years of Contributions to Virology. APS Press, St. Paul, MN.
3. Scholthof, K. B. 2004. Tobacco mosaic virus: a model system for plant biology. *Annu. Rev. Phytopathol.* 42:13–34.
4. Butler, P. J. 1999. Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:537–550.
5. Kegel, W. K., and P. van der Schoot. 2006. Physical regulation of the self-assembly of tobacco mosaic virus coat protein. *Biophys. J.* 91:1501–1512.
6. Klug, A. 1999. The tobacco mosaic virus particle: structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:531–535.
7. Caspar, D. L., and K. Namba. 1990. Switching in the self-assembly of tobacco mosaic virus. *Adv. Biophys.* 26:157–185.
8. Butler, P. J. G., and A. Klug. 1971. Assembly of the particle of tobacco mosaic virus from RNA and disks of protein. *Nat. New Biol.* 229:47–50.
9. Butler, P. J., J. T. Finch, and D. Zimmern. 1977. Configuration of tobacco mosaic virus, RNA during virus assembly. *Nature*. 265:217–219.
10. Butler, P. J. 1984. The current picture of the structure and assembly of tobacco mosaic virus. *J. Gen. Virol.* 65:253–279.
11. Shalaby, R. A., and M. A. Lauffer. 1977. Hydrogen ion uptake upon tobacco mosaic virus protein polymerization. *J. Mol. Biol.* 116:709–725.
12. Namba, K., and G. Stubbs. 1986. Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science*. 231:1401–1406.
13. Raghavendra, K., J. A. Kelly, ..., T. M. Schuster. 1988. Structure and function of disk aggregates of the coat protein of tobacco mosaic virus. *Biochemistry*. 27:7583–7588.
14. Fukuda, M., T. Ohno, ..., I. Takebe. 1978. Kinetics of biphasic reconstitution of tobacco mosaic virus in vitro. *Proc. Natl. Acad. Sci. USA*. 75:1727–1730.
15. Lomonosoff, G. P., and P. J. Butler. 1980. Assembly of tobacco mosaic virus: elongation towards the 3'-hydroxyl terminus of the RNA. *FEBS Lett.* 113:271–274.
16. Lebeurier, G., A. Nicolaieff, and K. E. Richards. 1977. Inside-out model for self-assembly of tobacco mosaic virus. *Proc. Natl. Acad. Sci. USA*. 74:149–153.
17. Shire, S. J., J. J. Stegkert, and T. M. Schuster. 1981. Mechanism of tobacco mosaic virus assembly: incorporation of 4S and 20S protein at pH 7.0 and 20°C. *Proc. Natl. Acad. Sci. USA*. 78:256–260.
18. Schuster, T. M., R. B. Scheele, ..., M. Potschka. 1980. Studies on the mechanism of assembly of tobacco mosaic virus. *Biophys. J.* 32:313–329.
19. Fukuda, M., and Y. Okada. 1985. Elongation in the major direction of tobacco mosaic virus assembly. *Proc. Natl. Acad. Sci. USA*. 82:3631–3634.
20. Caspar, D. L. 1980. Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* 32:103–138.
21. Erickson, J. W., and J. B. Bancroft. 1981. Melting of viral RNA by coat protein: assembly strategies for elongated plant viruses. *Virology*. 108:235–240.
22. Zlotnick, A., and S. Mukhopadhyay. 2011. Virus assembly, allostery and antivirals. *Trends Microbiol.* 19:14–23.
23. Zlotnick, A., J. M. Johnson, ..., D. Endres. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry*. 38:14644–14652.
24. Liepold, L. O., J. Revis, ..., T. Douglas. 2005. Structural transitions in Cowpea chlorotic mottle virus (CCMV). *Phys. Biol.* 2:S166–S172.
25. Dykeman, E. C., and R. Twarock. 2010. All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein. *Phys. Rev. E*. 81:031908.
26. Jabbari-Farouji, S., and P. van der Schoot. 2010. Competing templated and self-assembly in supramolecular polymers. *Macromolecules*. 43:5833–5844.
27. Janssen, P. G. A., S. Jabbari-Farouji, ..., A. P. Schenning. 2009. Insights into templated supramolecular polymerization: binding of naphthalene derivatives to ssDNA templates of different lengths. *J. Am. Chem. Soc.* 131:1222–1231.
28. Kittel, C. 1969. Phase transition of a molecular zipper. *Am. J. Phys.* 37:917–920.
29. Gallie, D. R., K. A. Plaskitt, and T. M. Wilson. 1987. The effect of multiple dispersed copies of the origin-of-assembly sequence from

- TMV RNA on the morphology of pseudovirus particles assembled in vitro. *Virology*. 158:473–476.
30. McGhee, J. D., and P. H. von Hippel. 1974. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86:469–489.
  31. Zimm, B. H., and J. K. Bragg. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31:526–535.
  32. Jernigan, R., J. Ferretti, and G. Weiss. 1973. Helix lifetimes within the conformational transition region. A random walk model. *Macromolecules*. 6:684–687.
  33. Brooks, C. L. 1996. Helix-coil kinetics: folding time scales for helical peptides from a sequential kinetic model. *J. Phys. Chem.* 100:2546–2549.
  34. Sano, Y., H. Inoue, ..., K. Kajiwara. 1995. Solution x-ray scattering study of reconstitution process of tobacco mosaic virus particle using low-temperature quenching. *Biophys. Chem.* 55:239–245.
  35. Sano, Y., H. Inoue, and Y. Hiragi. 1999. Differences of reconstitution process between tobacco mosaic virus and cucumber green mottle mosaic virus by synchrotron small angle x-ray scattering using low-temperature quenching. *J. Protein Chem.* 18:801–805.
  36. Endres, D., and A. Zlotnick. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* 83:1217–1230.
  37. Morozov, A. Y., R. F. Bruinsma, and J. Rudnick. 2009. Assembly of viruses and the pseudo-law of mass action. *J. Chem. Phys.* 131:155101.
  38. Hagan, M. F., and O. M. Elrad. 2010. Understanding the concentration dependence of viral capsid assembly kinetics—the origin of the lag time and identifying the critical nucleus size. *Biophys. J.* 98:1065–1074.
  39. Wu, X., Z. Xu, and J. G. Shaw. 1994. Uncoating of tobacco mosaic virus RNA in protoplasts. *Virology*. 200:256–262.
  40. Wu, X., and J. Shaw. 1996. Bidirectional uncoating of the genomic RNA of a helical virus. *Proc. Natl. Acad. Sci. USA*. 93:2981–2984.
  41. Shaw, J. G. 1999. Tobacco mosaic virus and the study of early events in virus infections. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:603–611.
  42. Mundry, K. W., P. A. Watkins, ..., T. M. Wilson. 1991. Complete uncoating of the 5' leader sequence of tobacco mosaic virus RNA occurs rapidly and is required to initiate cotranslational virus disassembly in vitro. *J. Gen. Virol.* 72:769–777.
  43. Culver, J. N. 2002. Tobacco mosaic virus assembly and disassembly: determinants in pathogenicity and resistance. *Annu. Rev. Phytopathol.* 40:287–308.
  44. Butler, P. J., and J. T. Finch. 1973. Structures and roles of the polymorphic forms of tobacco mosaic virus protein. VII. Lengths of the growing rods during assembly into nucleoprotein with the viral RNA. *J. Mol. Biol.* 78:637–649.
  45. Schön, A., and K. W. Mundry. 1984. Coordinated two-disk nucleation, growth and properties, of virus-like particles assembled from tobacco-mosaic-virus capsid protein with poly(A) or oligo(A) of different length. *Eur. J. Biochem.* 140:119–127.
  46. van der Schoot, P. 1992. Structure factor of a semidilute solution of polydisperse rodlike macromolecules. *Macromolecules*. 25:2923–2927.
  47. Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, NY.
  48. Zandi, R., P. van der Schoot, ..., H. Reiss. 2006. Classical nucleation theory of virus capsids. *Biophys. J.* 90:1939–1948.