Kinetics of viral self-assembly: Role of the single-stranded RNA antenna

Tao Hu and B. I. Shklovskii
Theoretical Physics Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA
(Received 5 December 2006; revised manuscript received 2 February 2007; published 1 May 2007)

Many viruses self-assemble from a large number of identical capsid proteins with long flexible N-terminal tails and single-stranded (ss) RNA. We study the role of the strong Coulomb interaction of positive N-terminal tails with ssRNA in the kinetics of *in vitro* virus self-assembly. Capsid proteins stick to the unassembled chain of ssRNA (which we call an "antenna") and slide on it toward the assembly site. We show that at excess of capsid proteins such one-dimensional diffusion accelerates self-assembly more than ten times. On the other hand at excess of ssRNA, the antenna slows self-assembly down. Several experiments are proposed to verify the role of the ssRNA antenna.

DOI: 10.1103/PhysRevE.75.051901 PACS number(s): 87.15.Nn, 81.16.Dn, 36.20.−r

Viruses self-assemble in host cells from identical capsid proteins (CPs) and their genome which in many cases is a long single-stranded (ss) RNA molecule. Icosahedral viruses are formed from $60T$ CPs for only certain triangulation numbers T such as 1, 3, 4, or 7, etc. [1]. The Coulomb interaction between the CP and ssRNA plays an important role in their self-assembly [2–5]. Two recent papers [6,7] emphasized that CPs of a big class of $T=3$ and 4 viruses have long, flexible, N-terminal tails. They explored the role played in the energetics of the virus structure by the Coulomb interaction between the brush of positive N-terminal tails rooted at the inner surface of the capsid and the negative ssRNA molecule [see Fig. 1(a)]. It was shown [7] that virus particles are most stable when the total length of ssRNA is close to the total length of the tails. For such a structure the absolute value of the total (negative) charge of ssRNA is approximately two times larger than the charge of the capsid. This conclusion agrees with available structural data. (A similar result was obtained earlier [4] by assuming that the positive charge of the CP is smeared on the inner surface of the capsid.)

In this paper, we continue to deal with the electrostatic interaction of N-terminal tails and ssRNA, but switch our attention from thermodynamics to the kinetics of *in vitro* self-assembly. Most papers on *in vitro* kinetics study the self-assembly of an empty capsid at much higher than biological concentrations of salt, where the Coulomb repulsion of capsid proteins is screened and hydrophobic interactions dominate [8,9]. In Ref. [9] one can clearly discriminate the initial nucleation "lag phase," followed by the "growth phase," where the average mass of the assembled particles linearly grows with time. The recent study of the kinetics of self-assembly with the ssRNA genome emphasizes that CPs stick to ssRNA before the assembly [10,11], so that a virus is actually assembled from the linear CP-ssRNA complex. Not much is known about the nucleation and growth phases of such an assembly.

The goal of this paper is to understand the role of the large length of ssRNA in the kinetics of self-assembly at biological salt concentrations. We assume that after nucleation (for example, at one end of ssRNA) the capsid growth is limited by CP diffusion. We calculate the acceleration of self-assembly, which originates from the fact that, due to the Coulomb interaction of N-terminal tails with ssRNA, CPs stick to the ssRNA and slide on it to the assembly site. In this case, ssRNA plays the role of a large antenna capturing CPs from the solution and leading them to the assembly site. Figure 1(b) illustrates this process. We show below that for a $T=3$ virus this mechanism can accelerate self-assembly by approximately 15 times.

We consider a dilute solution of virus CPs with molecules of its ssRNA genome. For most of this paper we assume that the concentrations of the protein $c \sim 2Mc_R$, where c_R is the concentration of ssRNA and M is the number of proteins in the assembled virus (for $T=3$ viruses $M=60T=180$). In this case there are enough proteins in the system to assemble the virus around each ssRNA molecule, and c changes weakly in the course of assembly. Viruses, however, self-assemble only when the concentration c of CP is larger than some threshold concentration c_1 [4], which is similar to the critical micelle concentration for the self-assembly of surfactant molecules [12]. The critical concentration c_1 can be estimated as

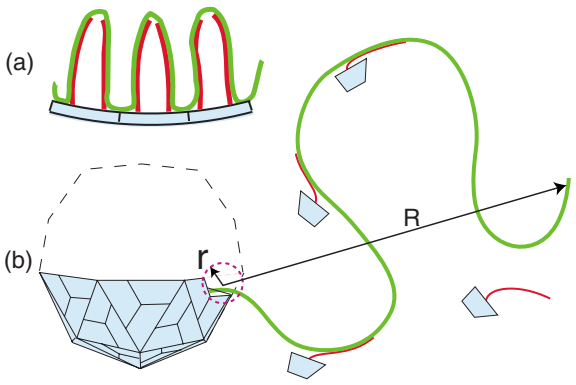

FIG. 1. (Color online) (a) An enlarged view from the inside of the virus. The brush of positive N-terminal tails [red (dark gray) line] is rooted at the inner surface of the capsid [blue (light gray) block]. The ssRNA [green (gray) line] strongly interacts with the tails and glues all the CPs together. (b) Schematic model of the capsid self-assembly. The unassembled ssRNA makes an antenna of size R for the one-dimensional pathway of the CPs toward the capsid assembly site at the capsid fragment (dashed circle with radius r of the size of a CP.

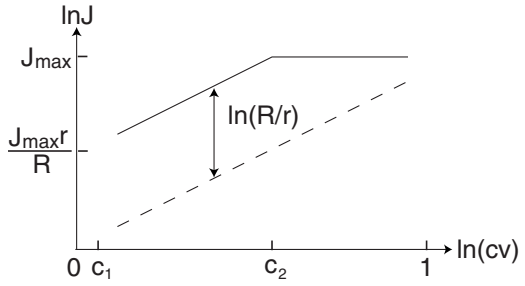

FIG. 2. Schematic plot of the diffusion-limited self-assembly rate J as a function of the protein concentration c. The full line is for the sliding of capsid proteins on ssRNA. The rate for the slower three-dimensional diffusion is shown by the dashed line.

$$c_1 \approx \frac{1}{v}\exp[-(\epsilon_e+\epsilon_p)/k_BT], \tag{1}$$

where v is the CP volume, ϵ_e is the absolute value of the electrostatic adsorption energy of the CP N-terminal tail to ssRNA, and ϵ_p is the absolute value of the CP-CP attraction energy in the capsid (per CP). Both ϵ_e and ϵ_p can be of the order of $10k_BT$, so that the critical concentration c_1 can be very small. In this paper we always assume that $c \gg c_1$. As shown in Ref. [7], in a partially assembled capsid, CP sticks to a piece of ssRNA of length equal to the tail length L [Fig. 1(a)]. A partially assembled capsid with $m<M$ CPs encapsulates the length mL of ssRNA. To continue this process the next $(m+1)$th CP should attach itself to the partially assembled capsid at the site, where ssRNA goes out of the capsid [see Fig. 1(b)] and this CP gets more nearest neighbors. We call this slowly moving site the "assembly site." It has a size of the order of the CP size r [see Fig. 1(b)].

CPs diffuse to the assembly site through bulk water. For $c \gg c_1$ one can neglect the dissociation flux from the assembly site. In this case the net rate of assembly (the number of CPs joining the capsid per unit time) is equal to the rate at which diffusing CPs find the absorbing sphere with radius r. It is equal to the Smoluchowski three-dimensional reaction rate [13]

$$J_3 = 4\pi D_3 r c, \tag{2}$$

where D_3 is the diffusion coefficient of the CP in water. The rate J_3 as a function of CP concentration c is plotted in Fig. 2 by the dashed straight line.

Our main idea is that the long chain of yet unassembled ssRNA outside the capsid provides an additional route for the diffusion of CPs to the assembly site, in analogy with the well-known faster-than-diffusion locating of the specific site on DNA for a protein [14,15]. The dramatic enhancement of the assembly rate is achieved because, due to the Boltzmann factor $\exp(\epsilon_e/k_BT)$, the three-dimensional concentration of CPs on the unassembled chain of ssRNA is larger than the bulk concentration c. This concentration can be estimated using a cylinder with cross section $v^{2/3}$ built around the ssRNA as the axis: it is equal to the number of CPs per unit length of ssRNA divided by $v^{2/3}$. At large distances the one-dimensional flux of CPs sliding on the ssRNA should be balanced by the three-dimensional diffusion flux of CPs to the ssRNA. This balance determines the radius ξ of the sphere around the assembly site at which two fluxes match each other and the crossover between three-dimensional and one-dimensional diffusion of CPs takes place. The ssRNA coil inside this radius is called an antenna.

The maximum possible antenna size is the characteristic size $R \sim (p\mathcal{L}_e)^{1/2}$ of the unassembled portion of ssRNA with the length $\mathcal{L}_e = \mathcal{L} - mL$. (Here we assume the ssRNA is a flexible Gaussian coil with the persistence length $p \sim 2b \sim 1.5$ nm, where $b \simeq 0.7$ nm is the monomer size, and we do not account for the excluded volume interaction.) In the case when $\xi = R$, the whole ssRNA adsorbs CPs arriving by three-dimensional diffusion and provides a path for fast one-dimensional diffusion to the assembly site (see Fig. 1). As a result, in this case the size R replaces the protein size r in Eq. (2) leading to a much faster rate

$$J = 4\pi D_3 R c, \tag{3}$$

which is shown in Fig. 2 by the part of the solid line parallel to the dashed one. Equation (3) is correct while CPs adsorbed on the unassembled chain of ssRNA are still sparse and do not block each other's diffusion on ssRNA. Let us use the notation c_2 for the concentration c, where the antenna becomes saturated by CPs and the dependence of the self-assembly rate J on c saturates roughly speaking at the level $J_{max} = 4\pi D_3 r/v$, which is the Smoluchowski rate J_3 at $c \sim 1/v$ (see the solid line in Fig. 2). It was shown in Ref. [15] that, if $\xi \leqslant R$,

$$c_2 = \frac{1}{v}\exp(-\epsilon_e/k_BT) \approx c_1 \exp(\epsilon_p/k_BT). \tag{4}$$

We see that the largest enhancement R/r of the self-assembly rate J can be achieved in the range of relatively small CP concentrations $c_1 \ll c \ll c_2$. For a typical $T=3$ virus the ssRNA genome consists of 3000 bases, so that the length $\mathcal{L} \sim 2100$ nm and $R \sim 60$ nm. Using $r \sim 4$ nm, we arrive at the acceleration factor $R/r \sim 15$. One can calculate the assembly time τ_a limited by diffusion. As we said above for $c \sim 2Mc_R$, the concentration of proteins c can be regarded as a constant. Thus, the antenna assisted assembly time τ_a is given by

$$\tau_a \approx \int_0^M \frac{dm}{4\pi c D_3[(M-m)Lp]^{1/2}} = \frac{2M^{1/2}}{4\pi c D_3 (Lp)^{1/2}}, \tag{5}$$

while, according to Eq. (2), the assembly time without the antenna is simply $\tau_0 = M/(4\pi c D_3 r)$. Since $(Lp)^{1/2} \sim 4$ nm, we can neglect the difference between $(Lp)^{1/2}$ and r, and arrive at the antenna assisted assembly time around $M^{1/2} \approx 14$ times shorter than τ_0

Strictly speaking, these estimates are correct only for self-assembly with a homopolymeric ssRNA or a synthetic negative polyelectrolyte [2]. For these cases, a small additional acceleration by a factor of 2 or 3 can be provided by the excluded volume effect. On the other hand, the native ssRNA is more compact than the Gaussian one due to the hydrogen bonds forming hairpins, and thus the estimated acceleration rate can be reduced by a factor of 2 or 3. Above, for simplicity we replaced ξ by its maximum value R. The

actual calculation of the antenna size ξ can follow the logic of the scaling estimate for the search rate of the specific site on DNA by a protein in Ref. [15]. In the current case, the assembly site plays the role of the target site (diffusion sink) for the protein, the unassembled chain of ssRNA plays the role of DNA, and the Coulomb attraction energy of the N-terminal tail to the unassembled ssRNA is analogous to the nonspecific binding energy of the diffusing protein on DNA. One may argue that the virus self-assembly problem is different, because the ssRNA plays a dual role. It is not only an antenna for the sliding CPs, but the ssRNA itself also moves to the assembly site, where it gets packed inside the capsid (each newly assembled CP consumes length L of the ssRNA). However, for a small concentration c in the range $c_1 \ll c \ll c_2$, where the unassembled ssRNA chain is weakly covered by CPs, the velocity of ssRNA drift in the direction of the assembly site is much smaller than the average velocity with which CPs drift along ssRNA. Thus, for the calculation of the assembly rate at a given length of the unassembled ssRNA chain we can use the approximation of a static ssRNA. This brings us back to the problem of proteins searching for the specific site on DNA [15]. Note that this means that the idea of self-assembly from the prepared linear ssRNA-protein complex [10,11] is literally correct only at $c > c_2$.

It is shown in Ref. [15] that, for a flexible ssRNA, the antenna size is $\xi \sim b(yd)^{1/3}$, where $y=\exp(\epsilon_e/k_BT)$ and $d=D_1/D_3$. D_1 is the one-dimensional diffusion coefficient of the protein sliding on ssRNA. This result remains correct as long as the antenna size ξ is smaller than the ssRNA coil size R. The adsorption energy ϵ_e of the N-terminal tail with approximately ten positive charges on the ssRNA can be as large as $10k_BT$. For $d=1$ we get $\xi \sim 30$ nm, while $R \sim 60$ nm. Thus, a simple estimate leads to the antenna length ξ somewhat smaller than R.

There are, however, two reasons why ξ may easily reach its maximum value R. First, some viruses self-assemble from dimers [9,11]. Naturally, dimers with their two positive tails bind to ssRNA with twice larger energy $2\epsilon_e$. This easily makes $\xi > R$. Second, the theory of Ref. [15] assumes that a protein has only one positive patch capable of binding to the double-helix DNA. Even if two segments of the DNA come in close contact, such a protein cannot simultaneously bind both segments and, therefore, cannot crawl between them without desorbing to water and losing the binding energy $-\epsilon_e$. For a globular protein this is a quite natural assumption. On the other hand, for a CP attached to the ssRNA by a flexible N-terminal tail, the tail can easily cross over (crawl) between the two adjacent segments of the same ssRNA, losing only a small fraction of the energy $-\epsilon_e$. This should speed up the protein diffusion on the ssRNA and may easily push ξ up to R.

Let us discuss ideas for three *in vitro* experiments, which can verify the role of the ssRNA antenna in the virus self-assembly. In the first experiment, one breaks the ssRNA into $K \gg 1$ short pieces of approximately equal lengths. It was shown [16,17] that the assembly is possible even when $K \sim M/2$, because, in order to glue CPs, the short ssRNA should bind two N-terminal tails of the neighboring proteins in the capsid. Virus assembly from short ssRNA pieces goes consecutively through two different diffusion-limited stages. In the first stage, capsid fragments (CFs) made of M/K proteins self-assemble on each short ssRNA molecule. According to Eq. (5), the time necessary for this stage is proportional to $(M/K)^{1/2}$ and is much shorter than the virus assembly time τ_a with the intact ssRNA. The second stage, where CFs aggregate to form the whole capsid takes much larger time τ_{as} (s stands for short). In order to calculate τ_{as} we assume that, when two CFs with n CPs each collide, they can relatively quickly rearrange their ssRNA molecules and CPs in order to make one bigger CF with $2n$ CPs. We also assume that at any time t all CFs have the same size $n(t)$. Then the concentration of such CFs is $c(n)=c_RM/n(t)$, where c_R is the concentration of original intact ssRNA. Therefore, the time required for doubling the number of CPs in a CF can be estimated from Eq. (2),

$$\tau(n) = \frac{1}{4\pi D_3(n) r(n) c(n)} = \frac{n}{4\pi D_3(n) r(n) c_R M}, \tag{6}$$

where $D_3(n)$ and $r(n)$ are the diffusion coefficient and effective radius of a CF with n CPs. Since the diffusion coefficient is inversely proportional to the droplet radius, the product $D_3(n)r(n)=k_BT/6\pi\eta$ (where η is the water viscosity) is the same constant as D_3r for a single protein. One collision of the droplets transfers n CPs to the growing CF. Therefore, the average time needed to add one CP to the growing CF $\tau_1=\tau(n)/n=1/4\pi D_3 r M c_R$ does not depend on n. In other words, the number $n(t)$ of CPs per CF increases at a constant rate. The assembly ends when n reaches M. Therefore, the assembly time is given by

$$\tau_{as} \simeq M\tau_1 \simeq \frac{1}{4\pi c_R D_3 r}. \tag{7}$$

Equation (7) shows that the assembly time depends on Mc_R, which stands for the concentration of CPs involved in the CF aggregation. However τ_{as} has no dependence on K. Comparing Eqs. (5) and (7), we obtain that at $c \sim 2Mc_R$

$$\frac{\tau_{as}}{\tau_a} \sim M^{1/2}\frac{(Lp)^{1/2}}{r} \sim M^{1/2} \gg 1. \tag{8}$$

We see that the virus assembly time with short ssRNA pieces is much larger than that for the intact ssRNA. This happens due to the breaking of big antenna of the original ssRNA.

In the second experiment, we return to the intact ssRNA and discuss what happens when we vary the relative concentrations of CPs and the ssRNA $x=c/Mc_R$, for example, keeping $c=$const and changing c_R. Until now we assumed that $x \sim 2$, i.e., we have marginally more proteins than is necessary to assemble a virus at every ssRNA. If $x \gg 1$ the assembly time τ_a is practically the same as that at $x \sim 2$ and is given by Eq. (5). Let us now consider much larger c_R, for which $x \ll 1$. Here the situation is dramatically different. There are two assembly stages. In the first stage, a CF is assembled with part of each ssRNA molecule, leaving the rest of the ssRNA molecule as a tail. This assembly uses up all the proteins and stops when all CFs are still much smaller than the complete capsid and their ssRNA tails are long [see, for example, Fig. 1(b)]. This state is essentially a kinetic trap.

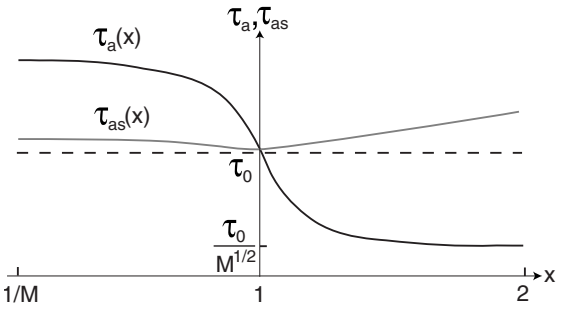

FIG. 3. Self-assembly times plotted schematically as a function of $x=c/Mc_R$. $\tau_0=M/(4\pi cD_3r)$ is the assembly time without the effect of ssRNA at $x>1$. The black and gray lines correspond to the intact ssRNA and short ssRNA pieces, respectively.

If the energies ϵ_e and ϵ_p are much larger than k_BT, CFs assembled on different ssRNA molecules cannot exchange CPs through the solution or via the collisions of their ssRNA tails. They can grow only via CF-CF collisions, which leads to their merging on one of the two involved ssRNA molecules and releasing the other empty one. We explained above, at $x>1$ (CPs are in excess), CFs without the ssRNA tails produce a capsid during time given by Eq. (7). On the other hand, at $x<1$, only the ssRNA molecules occupied by CPs take part in the aggregation. In order to calculate the assembly time, c_R in Eq. (7) should be replaced by c/M, which does not depend on x. However, due to the long ssRNA tail, a CF diffuses more slowly than it does without a tail. The time $\tau_a(x)$ grows substantially with decreasing x, because with more ssRNA, the initial CFs have fewer CPs and longer ssRNA tails. This time saturates at $x\sim 1/M$, where $c=c_R$ and each CF has only one protein and the longest ssRNA tail. Thus, a long antenna accelerates assembly at $x>1$ and decelerates it at $x<1$. This behavior of $\tau_a(x)$ is schematically plotted in Fig. 3.

In the third experiment we can combine the first two and break ssRNA into pieces at several different values of x. At $x<1$ the CF has a shorter tail of ssRNA and larger mobility, so that the assembly is faster than that for the intact ssRNA. When $x>1$, according to Eq. (7), the assembly time grows with decreasing c_R (increasing x). This is because, the smaller the ssRNA concentration, the harder for the CFs to collide with each other and form larger CFs. In other words, kinetics is determined only by CPs already assembled in CFs and their number decreases with growing x. We illustrate such nontrivial role of broken ssRNA in Fig. 3.

Now let us give some numerical estimates for c_1, c_2, and τ_0 for the *in vitro* assembly. Using the radius of a CP $r\sim 4$ nm, we obtain $c_1\sim 0.1$ nM and $c_2\sim 1$ μM from Eqs. (1) and (4). For $c\sim 1$ nM and the diffusion coefficient $D_3\sim 2\times 10^{-7}$ cm^2/s, the assembly time τ_0 is about 10 min. At excess of CP, ssRNA antenna reduces it to $\tau_a\sim 1$ min. At excess of ssRNA, roughly speaking τ_a increases to $2\tau_0$. One can make τ_a even larger using much longer than native ssRNA.

In conclusion, we studied the role played by the unassembled tail of ssRNA, which we call an antenna. We showed that the one-dimensional diffusion accelerates the virus self-assembly more than ten times when the CPs are in excess with respect to the ssRNA. On the other hand when the ssRNA molecules are in excess, the long ssRNA tail slows down the assembly. We discussed several experiments that can verify the role of the antenna. Although in this paper we focus on viruses whose CPs have long positive N-terminal tails, our idea can also be applied to the case where a CP can bind to the ssRNA by its positive patch. Our ideas are applicable beyond the icosahedral viruses, for example, to the assembly of immature retroviruses such as RSV or HIV [10,16,17].

We are grateful to V. A. Belyi, R. Bruinsma, A. Yu. Grosberg, T. T. Nguyen, A. Rein, Iu. Rouzina, V. M. Vogt, and A. Zlotnick for useful discussions.

[1] S. J. Flint, L. W. Enquist, V. R. Racaniello, and A. M. Skalka, *Principles of Virology*, 2nd ed. (ASM Press, Washington, DC, 2004).
[2] J. B. Brancroft, E. Heibert, and C. E. Bracker, Virology **39**, 924 (1969).
[3] K. W. Adolph and P. J. Butler, Philos. Trans. R. Soc. London, Ser. B **276**, 113 (1976).
[4] P. van der Schoot and R. Bruinsma, Phys. Rev. E **71**, 061928 (2005).
[5] R. F. Bruinsma, Eur. Phys. J. E **19**, 303 (2006).
[6] V. A. Belyi and M. Muthukumar, Proc. Natl. Acad. Sci. U.S.A. **103**, 17174 (2006).
[7] Tao Hu, Rui Zhang, and B. I. Shklovskii, e-print ArXiv: q-bio.BM/0610009.
[8] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, Virology **277**, 450 (2000).
[9] G. L. Casini, D. Graham, D. Heine, R. L. Garsea, and D. T. Wu, Virology **325**, 320 (2004).
[10] S. Campbell and V. M. Vogt, J. Virol. **69**, 6487 (1995).
[11] J. M. Johnson, D. A. Willis, M. J. Young, and A. Zlotnick, J. Mol. Biol. **335**, 455 (2004).
[12] D. F. Evans and H. Wennerstrom, *The Colloidal Domain*, 2nd ed. (Wiley-VCH, New York, 1999).
[13] M. V. Smoluchowski, Z. Phys. Chem., Stoechiom. Verwandtschaftsl. **92**, 129 (1917).
[14] O. G. Berg, R. B. Winter and P. H. von Hippel, Biochemistry **20**, 6929 (1981); R. F. Bruinsma, Physica A **313**, 211 (2002); S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2004); M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[15] Tao Hu, A. Yu. Grosberg, and B. I. Shklovskii, Biophys. J. **90**, 2731 (2006).
[16] S. Campbell and A. Rein, J. Virol. **73**, 2270 (1999).
[17] M. Y. Ma and V. M. Vogt, J. Virol. **76**, 5452 (2002); **78**, 52 (2004).