Polymorphism in Core-Controlled Virus Self-Assembly: Thermodynamics vs Kinetics

Orfeas-Agis Karachalios[⋆], Roya Zandi[‡], and Paul van der Schoot[⋆ †]
[⋆] *Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands*
[†] *Technische Universiteit Eindhoven, Postbus 513, 5600 HB Eindhoven, The Netherlands*
[‡] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
(Dated: July 2014)

Motivated by recent experimental studies we investigate theoretically the phenomenon of polymorphism in core-controlled self-assembly of virus-like particles, where identical nanoparticles are encapsulated in differently sized shells formed by virus coat proteins. Our purpose is to understand how protein concentration, stoichiometry and preferred curvature of the protein shells ("capsids") influence the prevalence of differently sized nanoparticle-protein complexes. We invoke equilibrium statistical theory and classical nucleation theory to study how kinetic trapping influences diagrams of state. Key ingredient is the spherical cap model describing the free energy penalty associated with the rim proteins that have fewer neighbours and hence fewer favourable contacts with other proteins in a shell. We find that the state diagrams that we calculate in and out of equilibrium differ significantly, and that kinetics favours the co-existence of polymorphs at large stoichiometric ratios and/or protein concentration.

I. INTRODUCTION

Virus assembly is a process that can be described to a large extent by simple thermodynamical principles [1–4]. Virus proteins self-assemble around genetic material forming closed structures, under ambient conditions of mild acidity and low ionic strength. In the absence of genetic material, they form empty capsids for high enough ionic strength [5, 6]. Specifically, a large number of virus species form spherical structures with icosahedral symmetry that are made up of a protein shell and a core, and are characterized by the Caspar-Klug (CK) T-numbers, indicating that the shell is build up from $60\,T$ protein subunits. Capsids that form will be largely characterized by a single T-number [6–8].

In addition, genetic material can be replaced with a variety of negatively charged cargo, ranging from non-native RNA, synthetic polyanions to functionalized gold nanoparticles, and will form capsids of non-native T-numbers in observable quantities [9–12]. A phase diagram has been constructed for Brome Mosaic Virus (BMV) capsid assembly in the presence of RNA [6], using the ionic strength and pH as parameters.

However, additional parameters may affect the phase diagram, such as the ratio of subunits to cores [13] called the stoichiometric ratio, as well as the appearance of different types of capsid assembled. Using theoretical studies to improve our understanding of the assembly mechanism and its dependence on ambient conditions are vital for developing applications. These range from viral drug development, that interfere with the assembly process, the use of enveloped virus shells as vectors for drug delivery and diagnostic medicine [14, 15], as well as their use as nano-reactors in nano-engineering applications [16].

The assembly process is driven largely by non-specific interactions and is similar in that respect to formation of micelles [4]. It has been established that proteins attract each other through hydrophobic interactions, that is, by burying hydrophobic surfaces between each other [3] and at the same time are repelled by electrostatic interactions due to the net positive charges they carry [17]. Additionally, specific interactions between carboxylate pairs, called Caspar pairs, form by intaking either protons or divalent cations that stabilize the protein complex [18].

The electrostatic repulsion between subunits is attenuated by increasing ionic strength or by introducing negatively charged cargo in the protein buffer, so that subunits adsorb on it and assemble into filled capsids. In order for the proteins to interact and bind reversibly to each other and their cargo [5], it has been suggested their conformation is changed [19]. The different conformations are realized by induced fitting mechanisms mediated through hinges located along the protein chain [20, 21]. These different conformations are intimately connected to the curvature and the T-number of the capsid and imply that the proteins assemble cooperatively.

In the case of empty capsids, it has been suggested that the balance of forces between hydrophobic attraction and electrostatic repulsion dictates which T-number is expressed in the end product of assembly [17]. This is motivated by experiments where cleaving a charged part of the protein sequence will result into higher efficiencies for T-numbers lower than the native T-number, although this may also occur because the cleaving renders the subunits unable to assume certain conformations found in the native T-number capsid [22]. Empty capsid assembly is in general monodisperse, an exception being Hepatitis B Virus (HBV) that has been found to express two T-numbers under native conditions [23].

In the case of filled capsids, attraction to the negatively charged core as well as its radius have to be also considered in order to determine the product T-number. It has been reported in experiment [24] and theory [25] that a minimum critical charge density must exist on the core for assembly to occur, independent of core size. Additionally, the core size controls the T-number that is assembled [26]. This is a result of complexation of protein N-terminal tails and the core charge that dictates the number of subunits that has to be chosen so the interaction between charges on the subunits and the core is optimal [27].

Thus, it is possible to control the size and symmetry of the final product, an important result for potential applications. However, a clear picture of the state diagram has not been drawn. Co-existence of different T-numbers has been reported [11, 12, 21, 26] but the dependence of it on experimental parameters remains unclear. Finally, the presence of empty capsids also has to be considered in the phase diagram, since they can limit the core-controlled assembly for high enough protein concentrations.

Experiments have been done using single-stranded RNA plant viruses, mainly Cowpea Chlorotic Mottle Virus (CCMV) and its close relative Brome Mosaic Virus (BMV), as well as enveloped viruses, in particular Hepatitis B Virus (HBV), which are relevant to applications in medicine, since they can be used as drug vectors in humans. The native T-number is $T = 3$ for CCMV and BMV and $T = 4$ for HBV. However, when presented with functionalized gold nanoparticles or RNA genome with different length than their native RNA, they encapsulate it forming non-native T-numbers [12].

In an experiment by the Dragnea group [26], CCMV coat protein assembled in $T = 1, 2, 3$ by varying the radius of the core-particle and it was reported that capsids of neighboring numbers co-existed. Similar experiments were carried out with HBV virus, where the stability of VLPs was tested by varying the size of the core [21]. In both cases, independent of whether RNA or gold cores were used, co-existence of different T-numbers was reported.

To explain available experimental data, the assembly pathway and its dependence on the template as well as on the ambient conditions has to be determined. The presence of a critical concentration, the sparsity of intermediates of assembly [28, 29] and the high efficiency of encapsulation imply that capsid assembly is akin to a phase transition and in particular micellization [4]. Empty capsid assembly has been well described by kinetic and thermodynamic models, where a model of stepwise assembly by Zlotnick describes rather well the the observed kinetics [23]. This predicts the presence of a critical nucleus, the rate of formation of which limits assembly.

Hagan explores the results of a kinetic model for core-controlled assembly following the example of Zlotnick for empty capsids [25]. A simpler model for kinetics could be useful in order to study polymorphism in core-controlled capsid assembly over a larger range of system parameters. This would help design experiments in a more systematic fashion, demonstrating the importance of each experimental parameter and how it is expected to change the outcome of the experiment.

Classical Nucleation Theory (CNT) has a long history in the description of first-order phase transitions and has successfully described the relation between the nucleation rate and supersaturation for gas-liquid and liquid-solid transitions [30]. It is used to calculate the rate of formation of the critical nucleus that represents the rate-limiting step in the formation of the new phase and has been applied to the assembly of empty capsids [31].

Additionally, CNT provides a simple relationship between the driving force and the critical concentration for assembly that in turn depends on the ambient conditions of acidity and salinity, and allows us to define the quench depth as a function of subunit concentration and stoichiometric ratio of the reactants. The barrier to nucleation is represented by an effective rim tension of a spherical cap.

It has been shown by Luque *et al.* and Siber *et al.* that this is a good description for the assembly of Lennard-Jones particles constrained on a spherical surface [32, 33]. The use of an effective rim tension renders the model as inherently cooperative, an essential feature of virus assembly, and discriminates between different T-numbers. Thus, we expect CNT to give a good qualitative picture of polymorphism based on a minimal number of parameters.

The remainder of this article is as follows. In the following Sec. II, we review equilibrium theory for a system with one product of assembly and intermediates using a Debye-type mean field theory that leads to a law of mass action [34]. The dependence of the encapsulation efficiency on the way experimental parameters are varied is demonstrated, in particular the stoichiometric ratio r, which is used for calculating the supersaturation used in CNT.

In Sec. III, IV and V the theory of kinetics is presented. In Sec. III, the expression we find for the nucleation rate is analyzed using CNT in conjunction with the spherical rim energy to describe the free energy barrier to nucleation and its impact on the nucleation rate. In Sec. IV, we continue the description of kinetics by using simple geometrical arguments to describe the attachment frequency of a subunit to a partial capsid that leads to different types of kinetics and discuss the dependence of the steady-state nucleation rate on system parameters. In Sec V, we complete the description of the theory of kinetics, by giving the explicit time dependence of efficiencies and providing a first example of competition between filled and empty capsids.

In Sec. VI, competition is studied using both equilibrium and kinetic theory. Specifically, assembly curves are calculated as a function of subunit concentration and stoichiometric ratio in the case that two T-numbers can form. Equilibrium theory and kinetics lead to different state diagrams for the system, the main result of this work. Finally, we discuss the implications of our results and an outlook in the last Sec. VII.

II. EQUILIBRIUM THEORY OF CORE-CONTROLLED VIRUS CAPSID ASSEMBLY

Whe first give a brief overview of earlier work done on the encapsulation of RNA as far as it carries over to that of nanoparticles by Zandi and van der Schoot[13], with emphasis given to parts that are relevant to nucleation and the state diagram of assembly. We assume a dilute solution of free protein subunits, monodisperse functionalized gold nanoparticle cores and partial capsids of size n at fixed temperature. Proteins adsorb on the cores through electrostatic attraction and bind with some free energy $h < 0$ in units of thermal k_BT to them. Each core offers q adsorption sites. Concomitantly, the proteins condense on the core surface with a free energy $g < 0$, producing a filled virus-like particle (VLP), gaining a total binding free energy $g + h$. We will ignore empty VLPs, as they do not effectively compete with the filled VLPs for large enough h, and only mention them again at the end of this section.

We study the system in the grand canonical ensemble, so that the chemical potentials of the subunits μ_s and cores μ_c are fixed. The grand potential density Ω, also expressed in units of thermal energy k_BT, is the sum of free energies of free proteins $x_s(\ln x_s - 1) + x_s\,\mu_s$, cores $x_c(\ln x_c - 1) + x_c\,\mu_c$ and intermediates $x_n\,(\ln x_n - 1) + g_n\,x_n + n\,x_s\,\mu_s$, where x_s, x_c, x_n are ensemble averages of the dimensionless concentrations, μ_s, μ_c are the corresponding chemical potentials and g_n is the net binding energy of a partial capsid of size n [31]. Since capsid assembly is a nucleated process, partial capsids experience an energy penalty in addition to the binding energy gain; actually it is the difference of the actual gain and that of what a complete capsid would have gained if it had size n. This penalty is effectively described by a

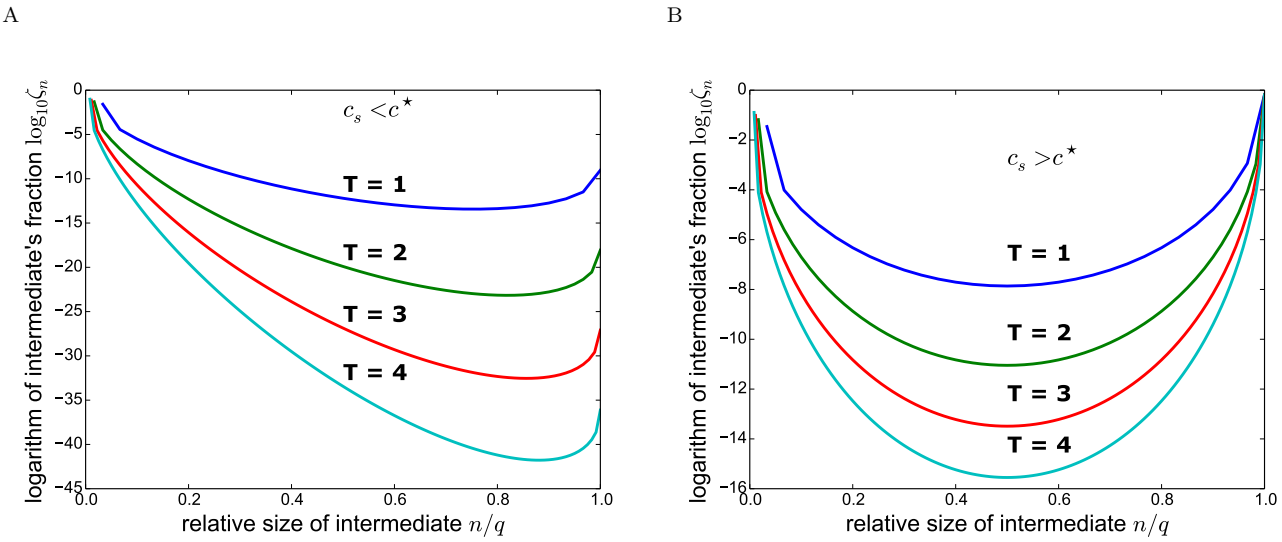


FIG. 1. Logarithmic fraction of partial capsids ζ_n as a function of their relative size n/q, where n is the number of subunits in the partial capsid and q is the number of subunits in the complete capsid. ζ_n is the ratio of concentration of intermediate of size n to the total concentration of cores c_c in solution. The concentration of subunits is $c_s = 0.5\,c^\star$ on A, where $c^\star$ is the critical concentration above which assembly occurs, and $c_s = 3\,c^\star$ on the right. The stoiciometric ratio $r = q\,c_s/c_c$ is equal to 1 and the ratio of critical concentration for empty capsid assembly to $c^\star$ is $d^\star/c^\star \simeq 50$. Each curve is for a different T-number, specifically $T = 1, 2, 3, 4$, which is connected to the size $q = 60 \times T$. Partial capsids are orders of magnitude less abundant than free subunits or the complete capsids, so that they can be effectively neglected from the free energy leading to an all-or-nothing model. This agrees with a rigorous result by Keef *et al.* [35], where intermediates of Papovaviridae virus assembly are found sparse predicted from tiling theory. This also motivates the use of Classical Nucleation Theory to model the kinetics of assembly, where intermediates are sparse.

rim tension σ due to a spherical rim of length with length l_n [31],

$$\sigma_n = \sigma\, l_n = \frac{4\pi\, R\, \sigma}{q}\sqrt{n\,(q-n)}, \quad \sigma = s\,\frac{|g|\sqrt{q}}{4R} \tag{1}$$

The rim tension σ depends only on the binding free energy between subunits. The geometric factor s is indicative of the bonds a subunit on the rim is missing, valued between 0 and 1. A study by Luque *et al.* that dealt with Lennard-Jones disks packed on a spherical surface found this to be $s \simeq 0.4$ [32]. This should be viewed as an upper limit to its value, since proteins are flexible compared to Lennard-Jonesian particles and can change their conformation to minimize line tension. In any case, the total binding free energy of a partial capsid is taken to be

$$g_n = n\,(g+h) + s\,\frac{\pi\,|g|}{\sqrt{q}}\sqrt{n\,(q-n)} \tag{2}$$

We assume s to be independent of T-number on the basis that the average number of bonds between units on the rim depends only on the subunit and not on the capsid geometry, while what changes is the strength of these bonds due to a non-optimal conformation. We note two interesting features of eqn. (2). Firstly, the total binding energy can be varied independently from the barrier height through h, thus changing the importance of the energy barrier and influencing the kinetics, which we will study in the next section. Secondly, the rim tension term does not depend explicitly on the core radius but depends only on the T-number as well as the binding energy per subunit, which is an implicit function of the T-number. This is expected, since the rim tension depends only on the geometry of the capsid itself and the properties of its subunits, so that it should be independent of the template.

Having defined the binding free energy g_n, we can derive the law of mass action from minimizing the grand potential [13]. This set of equations, together with the rim energy g_n, implies that intermediates are extremely sparse compared to the complete capsid. Specifically, $x_n/x_{n+1} = e^{g_{n+1}-g_n}/x_s << 1$ except for $n = q$, which is illustrated in Fig. 1A, 1B.

The fact that intermediates are sparse, implies that virus assembly is akin to a first-order phase transition and thus CNT is a candidate for describing the kinetics of it, which we will do in the next section. Additionally, we can dramatically decrease the complexity of the problem by focusing on an all-or-nothing model for the thermodynamics.

Thus, we write down a free energy that includes only complete capsids, free subunits and cores that leads to a single mass-action law.

Since the chemical potentials cannot be easily controlled in an experimental setting, we take advantage of the fact that ensembles become equivalent in the thermodynamic limit, switch to the canonical ensemble and interpret x_s, x_c and x_q as the concentrations after equilibrium is established in a system with limited supply of subunits $c_s = x_s + q\,x_q$ and cores $c_c = x_c + x_q$. We define the stoichiometric ratio $r = q\,c_c/c_s$, and the reduced core concentration $\tilde{r} = q\,c_c/c^\star$ as new control parameters and the encapsulation efficiency $\zeta = x_q/c_c$, the fraction of cores encapsulated, as the measurable variable. Keeping either r or $\tilde{r}$ fixed correspond to different types of experiment that result into different assembly curves. This will be shown for assembly curves predicted from equilibrium theory and kinetics in this and the following sections. It proves the importance of the stoichiometric ratio r as an assembly parameter that should be explicitly controlled in experiments.

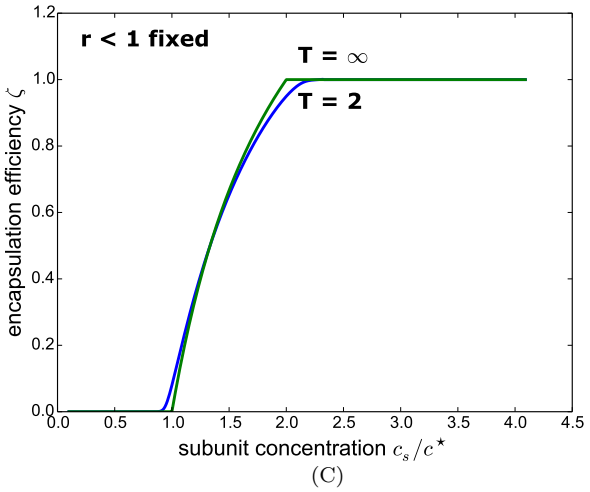
It is important to determine the encapsulation efficiency ζ as a function of protein subunit concentration c_s and stoichiometric ratio r, since the supersaturation relevant to CNT depends on it. Measurements of ζ are usually made for increasing c_s, however the dependence changes if we keep r or $\tilde{r}$ fixed. In the case that r is fixed, the solution is given by Zandi *et al.* [13] and is shown in Fig. 2A and Fig. 2B. Briefly, $\zeta \simeq 0$ for $c_s < c^\star$ and $\zeta \simeq (1 - c^\star/c_s)/\mathrm{r}$ above $c^\star$, where the critical capsid concentration $c^\star = e^{g+h}$ is the concentration below which no capsids assemble. For $r \geq 1$ capsid assembly is similar to empty capsid assembly, since cores do not limit the extent of assembly. For $r < 1$, the upper limit to the VLP population enhances assembly efficiency dramatically, which reaches 1 at a finite $c_s \simeq c^\star/(1-r)$. For the case where $\tilde{r}$ is fixed rather than r, $\zeta \simeq (c_s/c^\star - 1)/\tilde{r}$ so that the dependence is now linear as in Fig. 2C, which is observed in experiments [37]. Again, ζ reaches 1 at a finite $c_s = c^\star(1+\tilde{r})$.

In both cases, the maximum efficiency is achieved for subunit concentration c_s being a few times the critical concentration $c^\star$. This is important for kinetics, as we will see that the driving force of assembly depends on $(1/q)\ln\zeta$, which, for fixed $r < 1$ or fixed $\tilde{r}$, becomes practically zero over most of the range of subunit concentrations. This is because kinetics of assembly and therefore polymorphism depend on $\ln c_s/c^\star$ rather than $c_s/c^\star$, so that $c_s/c^\star$ has to vary over many orders of magnitude to study the state space of assembly. Thus, the efficiency of assembly is independent of the driving force of assembly within our model, leading to a simple expression for the driving force in the next Sec. III. Finally, the high efficiency of assembly and the all-or-nothing behavior motivates us to use the theory of first-order phase transitions to describe the kinetics of assembly.

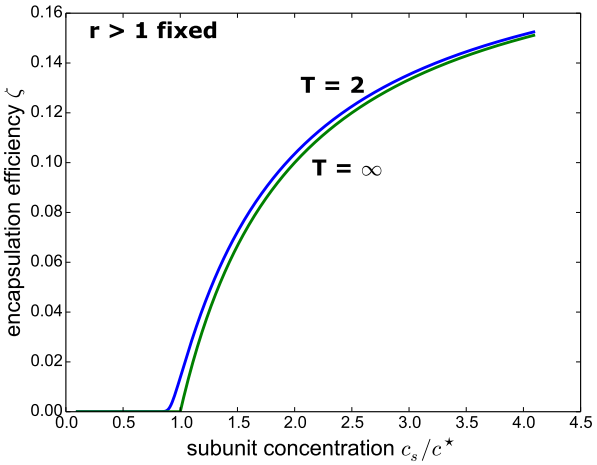
The theory by Zandi *et al.* generalizes to a system where two T-numbers form in the same solution and compete for both cores and protein [13]. We postpone the results for Sec. VI, so that we can first present the theory of kinetics in Sec. III,IV and V and make a direct comparison between them. Briefly, the ratio of the larger type to the smaller type is $x_{q_2}/x_{q_1} \sim (c_s/q_2 c_c)^{(q_2-q_1)/q_1}$, at least for large stoichiometric ratios r, so that it strongly depends on it. In contrast, the driving force of assembly does not, exactly because of the dependence on ζ described above.

The above leads to the first contradiction between equilibrium theory and kinetics, discussed in section VI and VII. We close this section with a state diagram of assembly of one type of capsid Fig. 3A, which is to be compared with the state diagrams in Sec. VI. The diagram allows for the formation of empty capsids above a critical concentration $d^\star$ for empty capsid assembly that is large compared to $c^\star$ and stoichiometric ratios $r < 1$, which does not change when two types of filled capsid are allowed to form. However, we will see in Sec. VI that kinetics give a different result, leading to a second contradiction. The theory of kinetics follows in Sec. III.

(A)



(B)



(C)

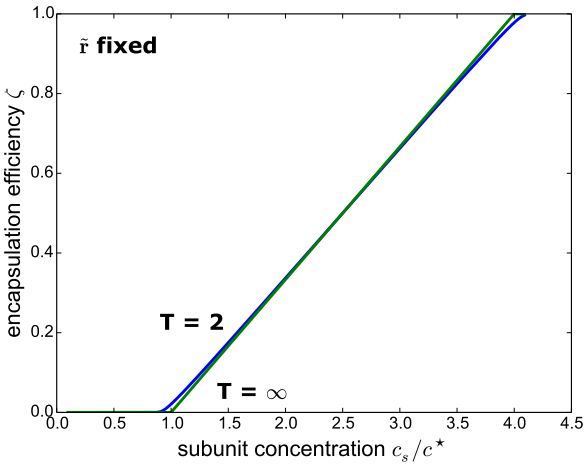


FIG. 2. Encapsulation efficiency ζ for mono-disperse assembly as a function of subunit (protein) concentration $c_s/c^\star$ as predicted from equilibrium theory, where $c^\star$ is the critical capsid concentration above which assembly occurs. The curves correspond to a $T = 2$ capsid and a capsid that has an infinite T-number. The two curves fall almost on top of each other, demonstrating that assembly is independent of capsid size as long as it is large enough. In all plots assembly starts at roughly $c^\star$, where a sharp rise in efficiency is observed. Two types of experiment are considered. Plots (A) and (B) correspond to an experiment with fixed stoichiometric ratio $r = q\,c_c/c_s$, where the concentration of cores c_c that assemble into capsids of $q = 60 \times T$ subunits is at a fixed ratio with c_s. For $r < 1$, specifically $r = 0.5$, assembly is extremely efficient, since ζ reaches 1 for c_s a few times $c^\star$. For $r > 1$, specifically $r = 5$, ζ is limited from above by $1/r$, here 0.2, and assembly converges to maximum efficiency at a slower rate with respect to c_s when compared to $r < 1$, but assembly is still efficient. In both plots, the dependence on c_s is non-linear, specifically $\sim c^\star/c_s$. Plot (C) corresponds to an experiment with fixed reduced core concentration $\tilde{r} = q\,c_c/c^\star$, here $\tilde{r} = 3$. The dependence on $c_s/c^\star$ is linear. Thus different types of experiment will result into different assembly curves, demonstrating the importance of the stoichiometric ratio as an assembly parameter.

(A)

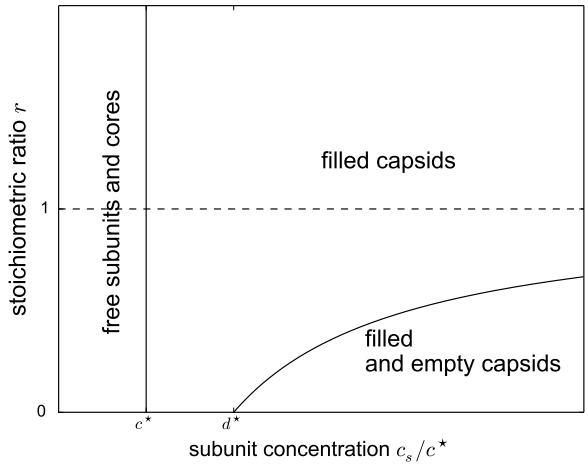


FIG. 3. Exact state diagram of mono-disperse core-controlled assembly as predicted from equilibrium theory, whose parameters are the subunit (protein) concentration c_s and stoichiometric ratio $r = q\,c_s/c_c$, where q is the number subunits in the capsid and c_c is the concentration of cores. Below the critical concentration for filled capsid assembly $c^\star$ there are only bare cores and free subunits, while above that most of the subunits are found in filled capsids. Empty capsids are also allowed to form above a critical concentration $d^\star$. For convenience, $d^\star$ is of a similar order of magnitude as $c^\star$, although this is not the case in experiment. If the stoichiometric ratio r is above 1, so that there are more cores than subunits, empty capsids cannot assemble at any c_s. Even if $r < 1$, the concentration of subunits after they are depleted to assemble capsids, $c_s - q\,c_v$, where c_v is the concentration of filled capsids, has to be larger than the critical concentration for empty capsid assembly $d^\star$. This leads to the appearance of a boundary between only filled capsids and co-existence of filled and empty capsids, whose form is $r = 1 - d^\star/c_s$. Thus, stoichiometric ratios above that value constitute disfavorable conditions for empty capsid assembly. However, empty capsids appear also at these stoichiometric ratios [36], a fact that may be explained by studying the kinetics of assembly.

III. NUCLEATION BARRIER

We study the kinetics of Core-Controlled Capsid Assembly using the framework of CNT. We expand on earlier work by Zandi *et al.* [31] for empty capsids, and assume that the work of formation of a partial capsid is given by

$$\Delta G(n) = -n\,\Delta\mu + \frac{4\pi\,R\sigma}{q}\sqrt{n\,(q-n)} \tag{3}$$

where the supersaturation $\Delta\mu$ is the chemical potential difference between the metastable free subunit state and the stable capsid state. This represents the so-called capillarity approximation, a central component of CNT [38]. As in the case of empty capsid assembly, single subunits attach to and detach from the nucleus reversibly and assembly is limited by the formation of the energetically unfavorable critical nucleus with size $n^\star$, which obeys $\partial\Delta G/\partial n = 0$.

We expect the presence of a template, represented by the core-particle, to change the assembly in three ways. Firstly, $c^\star$ depends on both the subunit-subunit interaction g and the protein-core interaction h, while the rim tension σ depends only on g, so that the critical concentration for capsid assembly can be changed independently from the barrier height. Secondly, the core acts as a nucleation center for the subunits so that the attachment rate to the nucleus is potentially enhanced compared to empty capsids. That is, apart from the free subunits in solution, there are also subunits adsorbed to the core of the capsid. Finally, the supersaturation $\Delta\mu = \ln(c_s/c^\star) - (\ln\zeta_{eq}(c_s, r))/q$ depends on the stoichiometric ratio r through ζ_{eq}, which is the encapsulation efficiency of assembly as predicted from the equilibrium theory in Sec. II. This is different from empty capsid assembly, which depends only on the concentration of proteins. We expand on these differences in the rest of the section.

Given the work of formation of the critical nucleus $\Delta G^\star = \Delta G(n^\star)$, the steady-state nucleation rate is given by the rate at which the critical nucleus forms [31]

$$J_S = c_c\,\nu^\star\,Z\exp\left(-\Delta G^\star\right) \tag{4}$$

where $\nu^\star$ is the attempt frequency of the critical nucleus to cross the nucleation barrier, $Z = \sqrt{-\partial^2\Delta G/\partial n^2|_{n^\star}/2\pi}$ is the Zeldovich factor, a measure of the flatness of the nucleation barrier and $\Delta G^\star = \Delta G(n^\star)$ represents the height of the nucleation barrier. An important feature of eqn. (4) is that the nucleation rate ranges over many orders of magnitude as the quench depth is varied. In particular, the rate is critically slow for shallow quenches and dominated by the nucleation barrier.

As the quench becomes deep, the pre-exponential factor becomes increasingly important, changing the dependence of the rate on assembly parameters, and in particular the size q. In order to study the kinetics of assembly, it is important to separate between of two regimes of the shallow and deep quenches and the contribution of each factor to the nucleation rate. We note that we treat g, h and q as independent parameters, even though we expect g and h to depend on the T-number, and thus q, in an experiment. Additionally, we introduce the ratio $d^\star/c^\star$ as a parameter of assembly, where $d^\star$ is the critical concentration for empty capsid assembly, which empty capsid has the same T-number as the filled capsid that assembles above $c^\star$.

We need a measure for the quench depth and to examine its dependence on c_s and r. Such a quantity is the supersaturation ratio $\Gamma = q\Delta\mu/4\pi R\sigma = \sqrt{q}\Delta\mu/\pi|g|s$ [31]. This is the ratio of the driving force of assembly $\Delta\mu$ to the barrier height $4\pi R\sigma/q$, and its small when the barrier is tall and large when the barrier is short. It is proportional to the supersaturation $\Delta\mu = \ln(c_s/c^\star) - (\ln\zeta_{eq})/q$, that depends strongly on c_s and weakly on r. If we choose to fix r, the last term is $(2.3/q)\,\log_{10}[r/(1-c^\star/c_s)]$, where $q \sim 100$, so that either the quench has to be extremely shallow, or $\log_{10} r \sim 50$, since $q \sim 100$. Fixing the reduced core concentration $\tilde{r}$ changes this to $(2.3/q)\,\log_{10}[\tilde{r}/(c_s/c^\star - 1)]$, so that the quench has to be extremely shallow. For deep quenches it leads to $\Delta\mu \sim (1-1/q)\,\ln(c_s/c^\star)$, which for large enough q is independent of the particular q.

Thus, the supersaturation is practically independent of r or $\tilde{r}$, depending on how the experiment is performed, for a large range of concentrations and the quench depth has a simple relationship to the subunit concentration, $\Delta\mu \simeq \ln c_s/c^\star$. This can be attributed to the large size q of virus capsids and the high efficiency of capsid assembly, so that the entropy of free cores is transformed into the entropy of VLPs and the only entropy cost results from the depletion of free subunits. We conclude that the rate does not have a discernible dependence on the stoichiometric ratio r or reduced core concentration $\tilde{r}$. However, the actual encapsulation efficiencies are determined from the system of rate equations, which depend on both the subunit and core concentration and therefore on r or $\tilde{r}$.

Having defined a measure for the quench depth, we can study the nucleation rate, beginning with the exponential Boltzmann factor. The critical nucleus work is rewritten in terms of the quantity Γ as $\Delta G^\star = (\sqrt{q}\,\pi\,s\,|g|/2)\left((1+\Gamma^2)^{1/2} - \Gamma\right)$ [31]. For shallow quenches the nucleation barrier governs the rate and $\Delta G^\star \simeq (\sqrt{q}\,\pi\,s\,|g|/2)(1-\Gamma) = \sqrt{q}\,\pi\,s\,|g|/2 - q\Delta\mu/2$. Larger capsids always have a higher barrier to nucleation at zero supersaturation but since $\Gamma \sim q$, the quench depth increases more strongly with the supersaturation compared to smaller capsids. Thus, there is a competition

(A) (B)

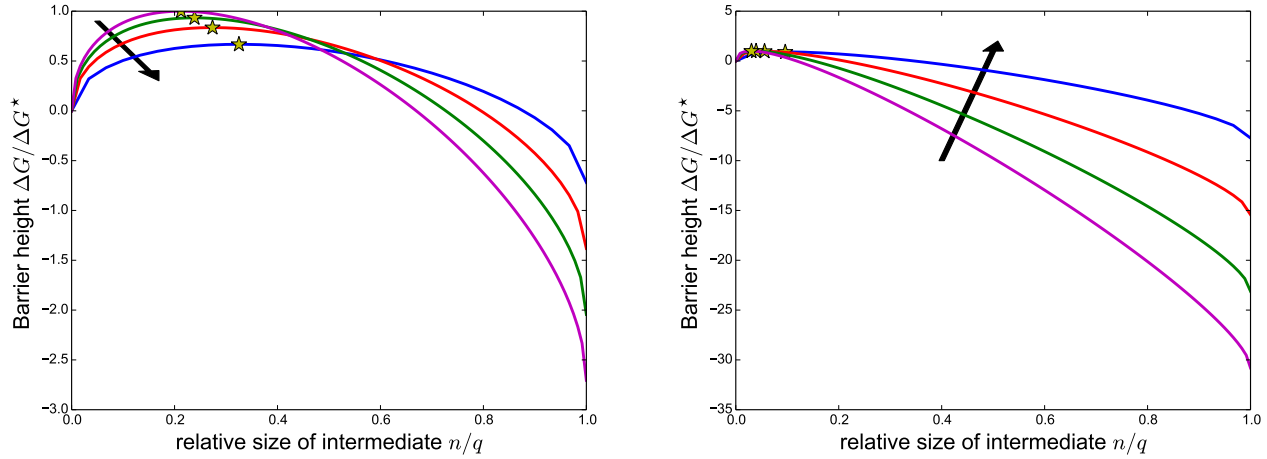


FIG. 4. The barrier to nucleation ΔG as a function of the relative size of intermediate n/q for different quench depths. Each curve is for a different T-number, which ranges from T=1 to 4. The barrier is normalized by the barrier height at the critical nucleus $\Delta G^\star$ of the largest capsid, $T = 4$. The order is indicated by the arrow and the T-number decreases along it. The star indicates the size of the critical nucleus, the most difficult to form intermediate. For shallow quenches, that is for subunit concentration $c_s \simeq c^\star$ and small Γ, the barriers have different height according to T-number, with the smaller T-numbers having shorter ones, and thus are easier to nucleate. We expect assembly to be monodisperse and the smallest capsids to form, at least if the critical concentration is the same for all T-numbers. For deep quenches, that is for $c_s >> c^\star$ and Γ large, the barriers level-off and have similar height. We expect all T-numbers to form at rates of similar order of magnitude and thus assembly should be polydisperse.

between the barrier height and the quench depth for increasing size q that determines which size is preferred for shallow quenches.

For mild to deep quenches, we have $\Delta G^\star \simeq \sqrt{q}\,\pi\, s\,|g|/4\Gamma = (\pi\, s\,|g|)^2/4\Delta\mu$, so that the exponential factor does not discriminate between different q, at least if g is equal for all q. Thus, we expect different T-numbers to have similar nucleation rates and therefore co-existence should be possible, at least when T-numbers have equal critical concentrations $c^\star$, which is an important prediction. However, the pre-exponential factor modifies this prediction, as it becomes significant compared to the exponential factor at deep quenches. Additionally, defining the pre-exponential factor requires that we define different types of kinetics. We study these different types and how they affect the nucleation rate in Sec. IV.

IV. STEADY STATE ATTACHMENT RATES

The nucleation rate is dictated also by the pre-exponential factor for deep quenches. It is the product of the Zeldovitch factor Z and the attempt frequency $\nu^\star$. The factor Z in particular represents the sharpness of the free energy barrier, so that the larger it is, the easier it is for the critical nucleus to cross the barrier. Alternatively, it can be seen as a measure of the lifetime of the critical nucleus, so as it becomes larger the critical nucleus survives for a longer period and has a greater chance to grow to a super-nucleus [30]. Its dependence on the assembly parameters follows from CNT and the form of the free energy landscape ΔG. In terms of the quantity Γ, $Z = (s\,g)^{1/2}(1/q + \Gamma^2/q)^{3/4}$ [31]. We see that $Z^{4/3} \sim 1/q + (\Delta\mu/\pi\,|g|\,s)^2$, so that smaller capsids experience a broader barrier.

The other part of the pre-exponential factor, $\nu^\star$, does not follow from CNT, but it depends on the geometry of the template and the subunits. It is the attempt frequency of a subunit to attach to the critical nucleus, thus promoting it to a super-nucleus. Since we are doing a qualitative analysis of capsid assembly, we write down an expression for $\nu^\star$ based on simple geometrical arguments. The attempt frequency is determined by the rate-limiting step in the attachment sequence of a subunit. Specifically, we envisage the attachment break down in three steps, shown in Fig. 5. Namely, diffusion and adsorption of the subunits to the core surface or rim, see Fig. 5A, followed by diffusion on the surface of the core towards the circular rim [Fig. 5B] and finally the conformational change of the subunit to become part of the capsid lattice [Fig. 5C]. We ignore direct attachment to the rim from the solution, since it competes directly with diffusion to the sphere and adsorption to the core surface. Since we are dealing with attachment to the

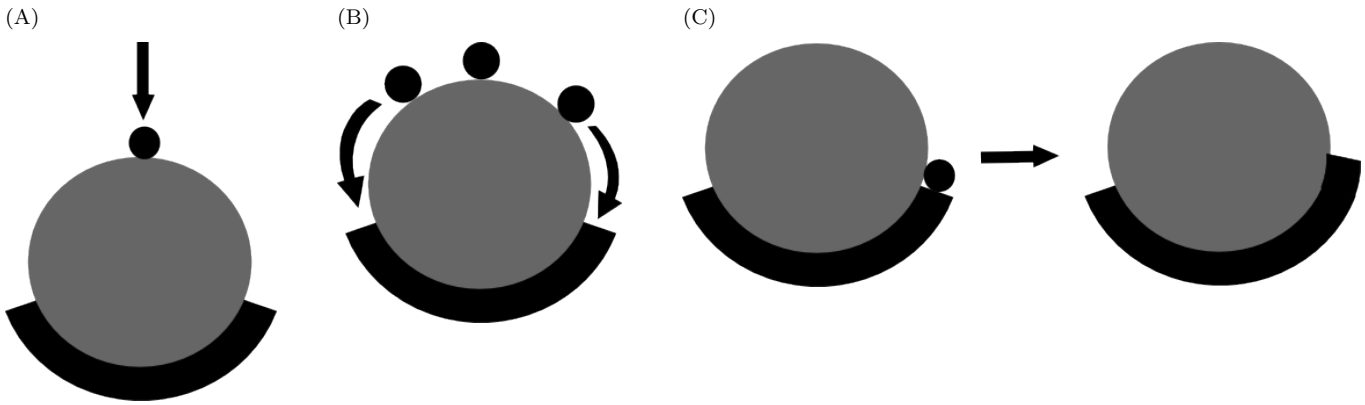


FIG. 5. The three modes of attachment, related to how a subunit attaches to the critical nucleus. In (A) assembly is diffusion-limited, meaning the subunit diffuses through the solution, adsorbs to the core and then immediately becomes part of the capsid. This is similar to stepwise assembly. In (B) assembly is surface-limited, where the subunit diffuses on the surface of the core towards the rim and then immediately becomes part of the capsid. This is similar to protein condensation. In (C) assembly is reaction-limited, where the subunit is located on the rim and has to change its conformation to become part of the capsid. This is also similar to protein condensation.

critical nucleus, the uncovered surface of the core is large compared to the rim, and attachment is more probable on the core. Plausibly, the rate is

$$1/\nu^\star = 1/\nu^\star_{diff} + 1/\nu^\star_{surf} + 1/\nu^\star_{reac} \tag{5}$$

so that the slowest step determines the effective attachment rate.

The rate for diffusion and adsorption to the core surface is proportional to the uncovered surface of the core

$$\nu^\star_{diff} = 2\pi\, R\, D\, c_s \left(1 + \frac{\Gamma}{(1+\Gamma^2)^{1/2}}\right) \tag{6}$$

where $R \sim 10\ nm$ is the radius of the sphere and D is the diffusion coefficient of proteins in aquatic solutions [26]. This follows from the solution of the diffusion equation for the subunit concentration towards a sphere with absorbing boundary conditions [30]. For mild to deep quenches, where the pre-exponential factor is important, the critical nucleus is small enough [31] that we can extrapolate from the formula for diffusion into a completely uncovered sphere and correct it with the Γ-dependent factor to account for any surface coverage. For deep quenches, this converges to a fixed value $\nu^\star_{diff} \simeq 4\pi\, R\, D\, c_s$, so that it depends only on the size of the core and is independent of q. Hence, the Zeldovitch factor Z determines the size dependence of the rate for deep quenches, which is $\sim (\Delta\mu/\pi\,|g|\,s)^{3/2}$ and the rate itself becomes size-independent and increases as $(c_s \ln c_s)^{3/2}$ for fixed r and $\sim c_s^2 (\ln c_s)^{3/2}$ for fixed $\tilde{r}$. It is then expected that co-existence of different sized VLPs will occur following deep quenches.

In contrast, the rate for the diffusion pathway on the core surface is proportional to the rim length, since subunits diffuse on the surface and stop as soon they reach the rim, which depends strongly on the size of the nucleus [31]. This gives

$$\nu^\star_{sur} = \pi\, R\, D\, c_s\, (1+\Gamma^2)^{-1/2} \tag{7}$$

which follows from the solution of the diffusion equation on the plane bound by a circular rim with absorbing boundary conditions and a constant incoming flux $D\, c_s/2$ equal to the rate of diffusion per area towards the sphere, that follows from the result for $\nu^\star_{diff}$. The concentration c_s is supposed to be smaller than $d^\star$, the critical concentration for empty capsid assembly, so that when the steady flux of particles on the surface has been set up, no depletion has occurred in bulk due to assembly of empty capsids. The dependence on Γ partially cancels the Zeldovitch factor, so that the pre-exponential factor is $\sim (\Delta\mu/q\pi\,|g|\,s)^{1/2}$. Consequently, the rate depends inversely on the size and smaller T-numbers nucleate faster. The rate increases more weakly with c_s, $\sim c_s\sqrt{\ln c_s}$ for fixed r and $c_s^2\sqrt{\ln c_s}$ for fixed $\tilde{r}$.

Finally, the rate of the third step should depend strictly only on the properties of the subunit and the interactions between subunits, so we set it equal to a constant $\nu^\star_{reac}$ independent of the quench depth, its order of magnitude being hard to estimate as it depends on the level of conformational switching the proteins undergo [41]. Comparing the different modes, we see that $\nu^\star_{diff}/\nu^\star_{surf} = 2\left((1+\Gamma)^{1/2} + \Gamma\right)$ which is always larger than 1. Thus the bulk diffusion

mode is always faster than surface diffusion mode within our model. For deep quenches, $\Gamma \gg 1$, it is much larger than surface diffusion. Therefore, the slowest mode dominates and we expect the kinetics in an experiment to be surface-limited, unless reaction-limited kinetics are extremely slow.

The steady state rate is plotted as a function of $\log c_s/c^\star$ for different sizes in Fig. 6A-6E. At shallow quenches assembly is monodisperse, since the rate changes orders of magnitudes with the size and only one phase will be able to form before subunits are depleted. At deep quenches the rate depends much more weakly on the size and different T-numbers are expected to form simultaneously, at least if binding energies do not depend on the size. Actually, simulations of virus assembly on charged polymers show this is not the case, as both the interaction with the core and between subunits depend on the T-number assembled [25, 27]. Even so, we will see in the Sec. VI that small differences in the critical concentration allow co-existence of different T-numbers.

We have to keep in mind, however, the nucleation rate depends strongly on the ratio of $d^\star/c^\star$, the critical concentration of empty capsid assembly to the one for filled capsid assembly, as it is illustrated in Fig. 6E. Keeping $c^\star$ fixed while varying $d^\star$ results into a different quench Γ, and this is reflected in the rate. Specifically, making the ratio larger increases the quench depth and the reverse holds as well. Thus, even if the total binding energies of different T-numbers are similar, the efficiencies ζ will be different, since for example a smaller capsid could have a stronger interaction with the core charge compared to a larger one if the template is commensurate with the smaller capsid radius. This can explain why only a limited range of T-numbers are observed in competition experiments.

This summarizes the dependence of the stationary rate of nucleation on assembly parameters. The steady-state rate alone does not determine the complete time dependence of encapsulation efficiencies and an additional assumption has to be made. In order to determine efficiencies, we define their time dependence explicitly in the next Sec. V and integrate them in Sec. VI.

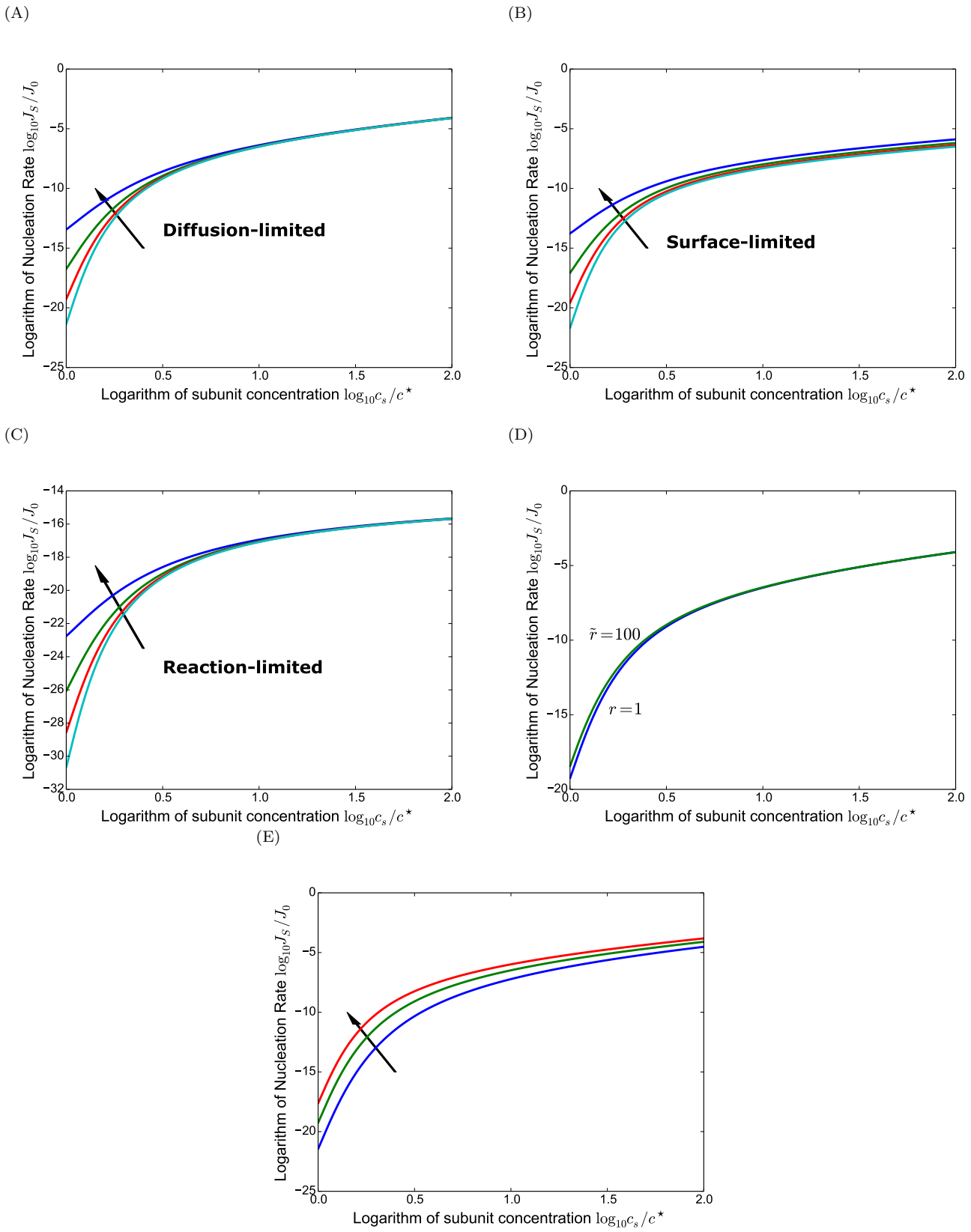


FIG. 6. Logarithm of the nucleation rate J_S as a function of the logarithm of subunit concentration $c_s/c^\star$, where $c^\star$ is the critical concentration above which assembly occurs. Rates differ many orders of magnitude at shallow quenches, that is $c_s \simeq c^\star$, so that we expect assembly to be monodisperse, favoring smaller capsids for equal $c^\star$ between T-numbers. At deep quenches rates are of the same order of magnitude and we expect T-numbers to co-exist. T-number ranges for T=1 to 4 and it decreases along the arrow. In figure (A) it can be seen that diffusion-limited kinetics do not differentiate between T-number at deep quenches. In contrast, surface-limited kinetics do differentiate, favoring smaller capsids. Reaction-limited kinetics, shown in figure (C), result into slower rates but they behave similarly to diffusion-limited kinetics at deep quenches. In figure (D), we compare the two types of experiments, where the stoichiometric ratio r or the reduced core concentration $\tilde{r}$ is kept fixed. J_S is not dependent on the stoichiometric ratio, as can be seen in the figure, for reasons explained in the text. Finally, in figure (E), we see the dependence on the ratio $d^\star/c^\star$, $d^\star$ being the critical concentration for empty capsid assembly, where we assume the empty capsid has the same T-number as the filled one. The arrow points in the direction of increasing ratio. The rate increases in the same direction, since the core-subunits interactions become large compared to inter-subunit interactions along it and the barrier to nucleation becomes smaller.

V. KINETICS OF CORE-CONTROLLED ASSEMBLY

The state diagram ultimately depends on the rates change of efficiencies with time. Within a zero-order approximation, common in applications of CNT [30], there are two phases of assembly. In the first phase the nucleation rate is zero for a time interval after application of the quench, called the time lag τ of nucleation. In the second phase, we assume the rate immediately reaches its stationary value from zero and protein subunits and cores start to deplete while VLPs are produced.

We first examine the first phase of assembly. The time lag of nucleation τ is predicted from CNT and is given in terms of the Zeldovitch factor Z and the attachment frequency $\nu^\star$

$$\tau = \frac{1}{4\pi^2\, Z^2\, \nu^\star} \tag{8}$$

so that the attachment mode has a large impact on the lag time. Diffusion and surface limited kinetics have a similar dependence on subunit concentration c_s, producing a time lag τ proportional to $(1/c_s)\,(1+\Gamma^2)^{-3/2}$ and to $(1/c_s)\,(1+\Gamma^2)^{-1}$ respectively, both being dominated by the factor $1/c_s$, so the two types of kinetics would result in similar lag times curves. On the other hand, reaction-limited kinetics have a much weaker concentration dependence with $\tau \sim (1+\Gamma^2)^{-3/2}$.

The lag time has been measured for empty Human Papilloma Virus (HPV) capsid assembly [39] and seems to follow the predictions of CNT for $\nu^\star \simeq constant$ (results not shown), but in the case of core-controlled capsid assembly the lag time seems instantaneous with respect to the assembly time and below the experimental limit for detection [40]. This is also reproduced by our model, as seen in Fig. 7A, where the interval during which the rate is zero is not noticeable. It is still interesting to study its dependence on the size and type of kinetics considered.

Lag times of different T-numbers do differ as seen in Fig. 7B-7D, with larger capsids having larger lag-times, however this does not influence the competition between T-numbers, since they are small compared to the time it takes for significant depletion of subunits and cores, at least within our model, even though we take them into account in the integration of the relevant evolution equations to be discussed next. Additionally, it is hard to determine the order of magnitude of reaction-limited lag-times compared to the other two types, as we need $\nu^\star_{reac}$, the rate at which a protein assumes the appropriate conformation when it reaches the rim. In contrast, diffusion and surface-limited kinetics are directly comparable, with diffusion-limited kinetics being an order of magnitude smaller.

After the lag time phase, capsids start being produced. We make the assumption that the system reacts fast enough to the subsequent depletion of protein subunits and cores by system of equations for the nucleation rates that depend on time due to the depletion of both protein subunits and cores,

$$\frac{d\,c_s}{dt} = -q\,J_S, \qquad \frac{d\,c_c}{dt} = -J_S, \qquad \frac{d\,c_v}{dt} = J_S. \tag{9}$$

with initial conditions, $c_s(0) = c_s^0$, $c_c(0) = c_c^0$ and $c_v(0) = 0$, which depend also on the stoichiometric ratio r. Hence, even though the rate J_S depend weakly on r, the solution of the above system of equations does depend on it.

In order to integrate eqn. 9, we assume a small time step $\Delta t = 0.001 c/J_S^0$, where c is the subunit concentration c_s for stoichiometric ratios $r > 1$ and the core concentration c_c for $r < 1$ and J_S^0 is the steady-state rate immediately after the lag time that is updated every time the rate decreases three orders of magnitude. These parameters where chosen in order for the integrated curves to be smooth and faithful to the actual solutions of eqn. 9. The concentrations are changed by finite differences and finally J_S is recalculated using the depleted values of c_s and c_c.

The integral of eqn. 9 is shown in Fig. 7A. We see that most of the subunits assemble in a short amount of time compared to the total time of assembly, at which point assembly becomes critically slow. This happens when the protein subunit concentration c_s approaches the critical concentration $c^\star$ from above, where the quench depth is extremely shallow and the barrier to nucleation dominates. This effectively means that almost no critical nuclei manage to assemble into capsids and the remaining subunits assemble into capsids at extremely large times compared to experimental times.

Finally, we compare the nucleation rate of capsids containing cores to empty ones by integrating a system of equations similar to eqn. 9 in the case where the subunit concentration $c_s > d^\star$ and assuming the empty capsids have the same size as the filled ones. As was mentioned earlier, the exponential factor in the expression for the empty capsid nucleation rate is the same as for core capsids, except that the $d^\star > c^\star$. That is, $\Delta G_e^\star$ has the same form as $\Delta G^\star$, except that the quench depth Γ depends on $d^\star$.

To completely define the nucleation rate for the empty capsid assembly, we also need an expression for the attachment rate to the critical nucleus. We envisage the attachment to an empty capsid to be given by direct attachment to the rim followed by a conformational change of the protein. The first step corresponds to diffusion-limited kinetics and we approximate it as the flux of subunits into a circular ribbon of width $b \sim 5nm$, an average width of a protein capsid

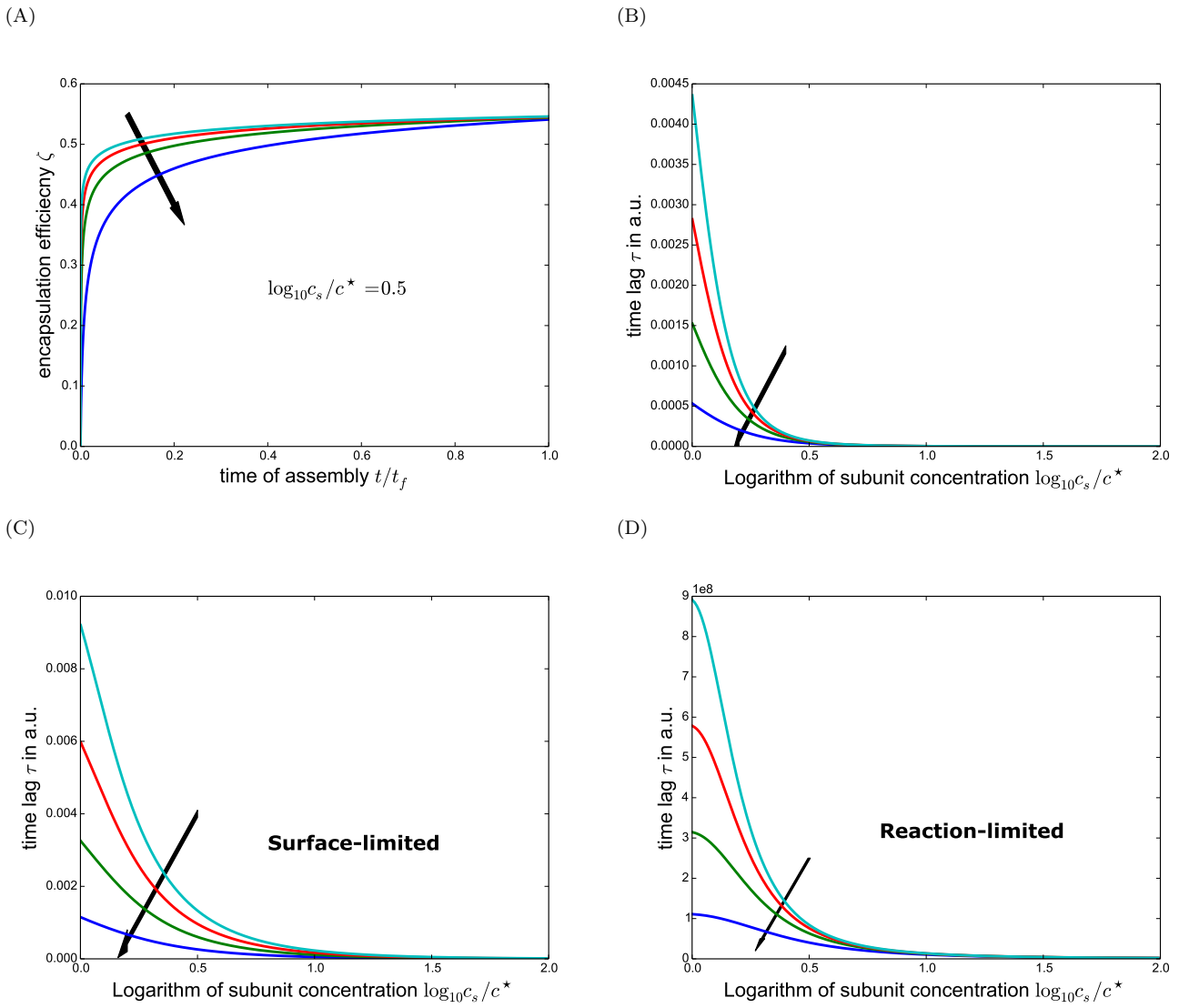


FIG. 7. Time dependence of the kinetics of assembly is shown for different T-numbers, fixed stoichiometric ratio $r = 1$. The arrow points the direction of decreasing T-number, from T=1 to 4. In plot (A) the explicit dependence of ζ is shown at a fixed quench depth $\log_{10} c_s/c^\star = 0.5$. Time is normalized to the time it takes for ζ to reach 0.8 of its maximum value predicted from equilibrium theory. Assembly is fast for small times and becomes critically slow as protein subunits are depleted. In plots (B), (C) and (D) time lags τ as predicted from CNT are plotted against $\log_{10} c_s/c^\star$ at a fixed stoichiometric ratio $r = 1$. The time units are $R\,D\,c^\star$, with R the core radius and D the diffusivity of proteins in solution, and are the same in all plots. Time lags differ between T-numbers only at shallow quenches, with larger T-numbers having larger time lags. The time lag is also dependent on the type of kinetics. Diffusion- and Surface-limited kinetics have a similar dependence on $\log_{10} c_s/c^\star$, although the latter predict larger lag times. Time lags are noticeable only for shallow quenches and become negligible fast. In contrast, reaction-limited kinetics predicts a constant time-lag for shallow quenches and display much larger time lags, although this depends strongly on the time it takes for a single subunit to change its conformation. Here we assume $\nu^\star_{reac} = 10^{-8} sec$, which is a mean time-scale for protein conformational changes [41].

and radius equal to the radius of the spherical rim $R\sin\theta^\star = R(1+\Gamma^2)^{-1/2}$, giving for the attachment frequency of the empty capsids

$$\nu_e^\star = 2\pi\, R\, D\, c_s (1+\Gamma_e^2)^{-1/2}\, f(b, R\sin\theta^\star) \tag{10}$$

where the function f follows from the solution of the diffusion equation in cylindrical coordinates and contains a zeroth order Bessel function, but is of secondary importance, as its argument should be small for deep quenches so that the attachment rate is important compared to the barrier, and therefore its value close to 1.

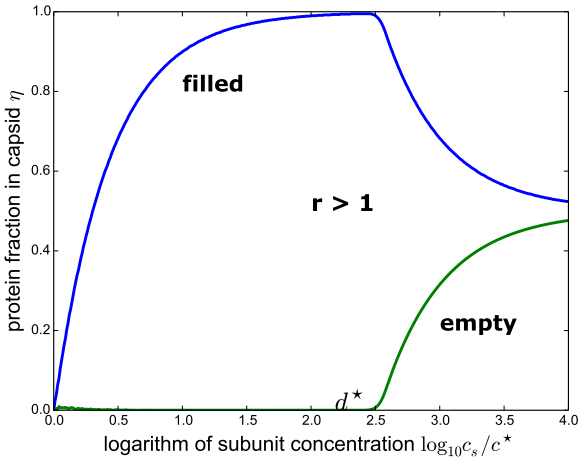


FIG. 8. Protein fraction in a capsid product $\eta = q\,c/c_s$, where c is the concentration of the capsid product, q the number of subunits in it and c_s the concentration of protein subunits, as a function of logarithm of subunit concentration $c_s/c^\star$, where $c^\star$ is the critical concentration for filled capsid assembly. This is predicted from surface-limited kinetics. Both filled and empty capsids assemble at an equal T-number, specifically T=3. This takes place for stoichiometric ratio $r = q * c_c/c_s > 1$, the ratio of core to subunit concentration, here $r = 5$. For $r \to 0$, most of the protein assembles into empty capsids. However, for $r > 1$, no empty capsids should assemble according to equilibrium theory. However, kinetics allows for the formation of empty capsids, since they form fast enough for deep enough quenches at a concentration above the critical concentration for empty capsid assembly $d^\star \simeq 100c^\star$. This does occur in experiment and it motivates the use of kinetic theory to predict the state diagram of core-controlled assembly [36].

Following Zandi *et al.* [31], the steady-state rate of formation of empty capsids has the form

$$J_{S,e} = c_s\,\nu_e^\star\,Z_e\,\exp(-\Delta G^\star(\Gamma_e)) \tag{11}$$

The quantity $\Gamma_e = \sqrt{q}\Delta\mu_e/(\pi|g|s)$ is the quench depth for empty capsids, where $\Delta\mu_e \simeq \ln c_s/d^\star$. To compare with filled capsids, we calculate the ratio $\Gamma_e/\Gamma \simeq 1 - 2/\log_{10} c_s/c^\star$, if $d^\star/c^\star \sim 100$. We see that the quench depth, which corresponds to empty capsids, is shallower compared to the one of filled ones at the same protein subunit concentration c_s. Empty capsids may form only for deep quenches, where the pre-exponential factor is important. In this case, the rate expressions for core and empty capsid assembly mainly differ in the fact that $J_{S,e} \sim c_s\,c_c$ and $J_{c_e} \sim c_s^2$, so that the ratio of rates is $\sim q/r$. Thus, we expect empty capsids to dominate over filled ones for stoichiometric ratio $r << q$ and for deep quenches. Additionally, if the reduced core concentration $\tilde{r}$ is fixed, the stoichiometric ratio $r \to 0$ with increasing protein concentration, so that empty capsids will dominate. Competition results of empty capsid assembly against surface-limited kinetics are shown in Fig. V.

It is important to point out that the competition between filled and empty capsids depends also on the assembly protocol. If cores and subunits are mixed in a solution with non-acidic pH and low ionic strength and then lowered to acidic pH as was done by Dragnea *et al.*, subunits will adsorb first on the cores and when pH is lowered, so that the system is quenched, the free subunit concentration for assembly may be below the critical concentration for assembly of empty capsids $d^\star$ [37]. In the following calculations, we consider only core VLPs form, assuming the above assembly protocol.

VI. POLYMORPHISM IN CORE-CONTROLLED CAPSID ASSEMBLY

As announced above, equilibrium theory and kinetics lead to a different prediction for the dependence of polymorphism on the assembly parameters, in particular the quench depth $\log c_s/c^\star$ and stoichiometric ratio r. In addition, kinetics is sensitive to the ratio of critical concentrations $d^\star/c^\star$. Here we study the case where two capsid types are allowed to form both for simplicity and because we expect binding energies to differ but as little as possible between the two types, namely T=2 and 3 capsids which are labeled 1 and 2 in the text and figures [27].

When two types of filled capsid compete, equilibrium theory predicts a strong dependence of polymorphism on the stoichiometric ratio r [13]. In particular, for equal critical concentrations between the two types, it is found that small capsids are favored at stoichiometric ratio r above 1, while large capsids dominate at strictly $r < 1$. There is a narrow range of r where the two types co-exist as seen in Fig. 9A.

The above dependence can be understood in terms of the system maximizing the free energy. For a stoichiometric ratio r larger than 1, the system maximizes entropy by distributing protein over the largest possible number of VLPs, which occurs if smaller capsids form. For $r < 1$, the system must pack as much protein on cores as possible in order to maximize the energy gain due to binding between proteins, so that large capsids form as in Fig. 9C. This behavior can be entirely seen in the assembly curve for fixed reduced concentration $\tilde{r} = 100$, Fig. 9D, where T=2 capsids form at shallow quenches, so that $r = \tilde{r}c^\star/c_s$ is smal, and T=3 form at deep quenches and r small. Again, the two types of experiment produce different assembly curves.

Finally, the two types also co-exist for shallow quenches and $r < 1$, as seen if Fig. 9C, where encapsulation efficiency of both types has not reached the maximum possible efficiency, that is, $\zeta_1 + \zeta_2 < 1$. The quench depth $\Gamma \sim \log_{10} c_s/c^\star$ has little impact on polymorphism otherwise. This is summarized in the state diagram in Fig. 12A. If we turn on a difference in the critical concentrations, favoring the large capsids, the small ones fail to compete at any r as in Fig. 9B and the phase diagram turns into the monodisperse one in Fig. 3A. The reverse occurs if the difference in critical concentrations are reversed too.

Compared to equilibrium theory described above, kinetics has a different dependence on both the quench depth $\sim \log_{10} c_s/c^\star$ and the stoichiometric ratio r, but is also sensitive to the attachment mode and the ratio of critical concentrations $d^\star/c^\star$ of empty capsid assembly to filled capsid assembly of the same T-number. Usually, a certain virus assembles into only one T-number when forming empty capsids, say T=3 for example, so $d_1^\star$ that corresponds to T=2 would not be able to be directly measured in a real experiment. However, it may be possible to cleave the protein sequence and allow for the formation the smaller capsid in order to measure $d_1^\star$, as was done by Zlotnick for HBV [22].

We assume that the protein subunit concentration is always $c_s < d^\star$, so that no empty capsids assemble. The time dependence for the efficiencies now take the form

$$\frac{d\,c_s}{dt} = -q_1\,J_{S,1} - q_2\,J_{S,2}, \qquad \frac{d\,c_c}{dt} = -J_{S,1} - J_{S,2}, \qquad \frac{d\,c_1}{dt} = J_{S,1}, \qquad \frac{d\,c_2}{dt} = J_{S,2} \tag{12}$$

with initial conditions $c_s(0) = c_s^0$, $c_c(0) = c_c^0$, $c_1(0) = 0$ and $c_2(0) = 0$. We integrate the equations as in Sec. V and include the time lags in the kinetic description. The results demonstrate how kinetics differs from equilibrium theory in its prediction.

Firstly, polymorphism depends more weakly on r, through the initial conditions and eqn. (12). In particular, the stoichiometric ratio does not impact polymorphism strongly, as the steady-state rate is very weakly dependent on it, see Sec. III. It does impact the sharpness of the assembly curve of the capsid type with the higher efficiency, when the two types reach maximum efficiency $\zeta_1 + \zeta_2 = 1$, as seen by comparing the case where $r < 1$ Fig. 10A and $r > 1$ Fig. 10C.

In comparison, polymorphism depends strongly on the quench depth $\Gamma \sim \log_{10} c_s/c^\star$. For shallow quenches assembly is monodisperse while at deep quenches the efficiencies of the two T-numbers are on a similar order of magnitude as seen in Fig. 10A, 10B and 10C and we have co-existence. This is as expected from Sec. IV, where steady-state rates differed orders of magnitude for shallow quenches and are similar for deep quenches. Fixing the reduced core concentration $\tilde{r}$ leads to a different assembly curve Fig. 11A. In the case that the critical concentrations $c_1^\star = c_2^\star$, for shallow quenches the small T-number is always preferred, since the barrier to nucleation corresponding to them is shorter.

If we turn on a difference between the critical concentrations $c_1^\star > c_2^\star$, the picture is reversed for shallow quenches and the large T-number forms Fig. 10D. The barrier to nucleation is still smaller for smaller capsids, however the quench depth $\Gamma \simeq 2.3\sqrt{q}\log_{10} c_s/c^\star/(\pi|g|s)$ increases more strongly with the protein concentration for the large T-number. In contrast to the results of equilibrium theory, the energy difference does not influence the co-existence of the system at deep quenches, since the difference between the critical concentrations of the two T-numbers becomes negligible as the quench depth is increased. Since $c_1^\star > c_2^\star$ in experiment, we expect co-existence to occur if kinetics are important in capsid assembly.

Co-existence of T-number for deep quenches is affected by the type of kinetics. Diffusion- and reaction-limited kinetics show no preference for any specific T-number and encapsulation efficiencies become equal with increasing protein subunit concentration c_s as seen in Fig. 10A. Surface-limited kinetics do differentiate between T-numbers and more T=2 capsids form than T=3, Fig. 10B. This follows from the attachment modes : the attachment rate is $\nu_{surf}^\star \simeq (1+\Gamma^2)^{-1/2}$ in surface-limited kinetics, so that the large T-number that is quenched more deeply at the same protein concentration as the small T-number, since $\Gamma \sim \sqrt{q}$, forms at a slower rate for deep quenches. This is not affected by a difference in critical concentrations, since again $\Gamma_2 > \Gamma_1$. Reaction-limited kinetics are not shown for brevity, as they are similar to diffusion-limited kinetics.

Finally, co-existence of T-numbers is influenced by the ratio $d^\star/c^\star$ of critical concentrations for empty capsid assembly to the one for filled capsid assembly. By making the ratio larger for the small capsids while keeping the critical concentrations $c_1^\star = c_2^\star$, effectively increasing the subunit-core interactions in the small capsids compared to

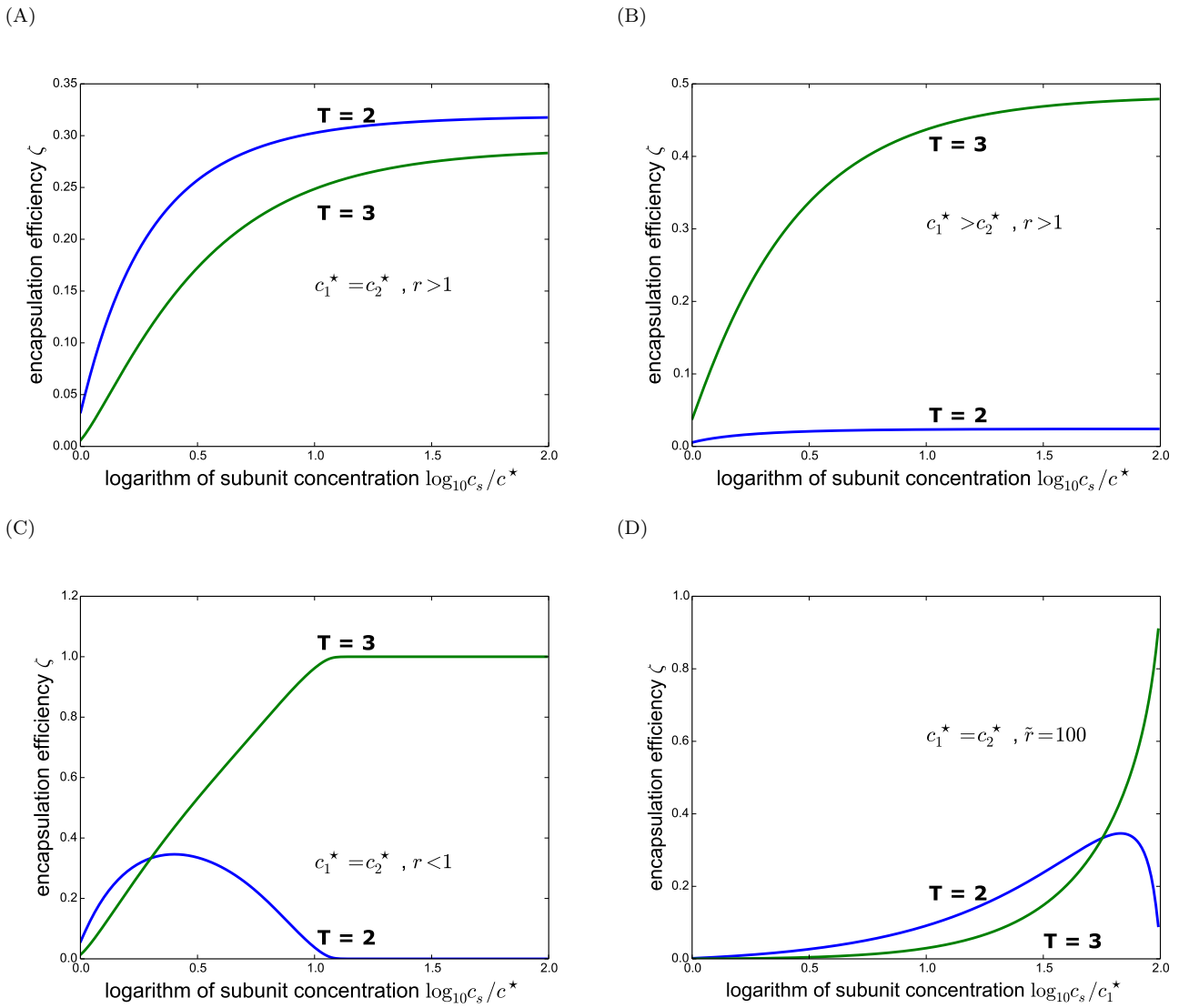


FIG. 9. Encapsulation efficiency ζ of filled capsids of T-number $T = 2$ and $T = 3$ that compete for subunits and cores as a function of the logarithm of subunit concentration $c_s/c^\star$, where $c^\star = c_1^\star$ is the critical concentration for T=2 capsid assembly, predicted from equilibrium theory. In figures (A) and (C) the dependence on the stoichiometric ratio $r = q_2\, c_c/c_s$ is emphasized, where q_2 is the number of subunits in a T=3 capsid and c_c is the core concentration. No T-number is energetically preferred, that is, $c_1^\star = c_2^\star$, where $c_2^\star$ is the critical concentration for T=3 capsid assembly. Specifically, for a narrow range of stoichiometric ratio r, above and close to 1, the two types co-exist at deep quenches $c_s \gg c_1^\star$, as in plot (A). If the stoichiometric ratio r is below 1, as in figure (C), the larger T-number is preferred and assembly is monodisperse at deep quenches. For stoichiometric ratios r large compared to 1 smaller capsids are preferred and assembly is again monodisperse. This is not shown here for reasons of brevity, but these results are summarized in Zandi *et al.* [13]. In figure (B), larger capsids are slightly preferred and $c_1^\star > c_2^\star$, which results into assembly being strictly monodisperse, favoring the larger capsid. This is in the region of stoichiometric ratio of figure (A). Thus, polymorphism in equilibrium theory is very sensitive to the ratio of critical concentrations of the two capsid types. Finally, in figure (D) the alternative type of assembly curve is depicted, where the reduced core concentration $\tilde{r} = q_2\, c_c/c_2^\star$ is kept fixed. Since c_s increases while c_c is fixed, the stoichiometric ratio $r = q\, c_c/c_s$ goes to zero and assembly is monodisperse, favoring larger capsids as in figure (C).

subunit-subunit interactions, the quench depth becomes much deeper for small capsids at the same protein subunit concentration c_s, because of the dependence $\log_{10} c_s/c_1^\star$. For deep quenches, small capsids tend to dominate as the ratio $d^\star/c^\star$ becomes larger Fig 11B. This may explain why only a limited set of T-numbers co-exist for deep quenches, as T-numbers that are incompatible with the core-template or that form at a curvature much different from the wild-type one are not quenched as deeply.

To summarize how kinetics modify the state space of capsid assembly compared to equilibrium theory, we present

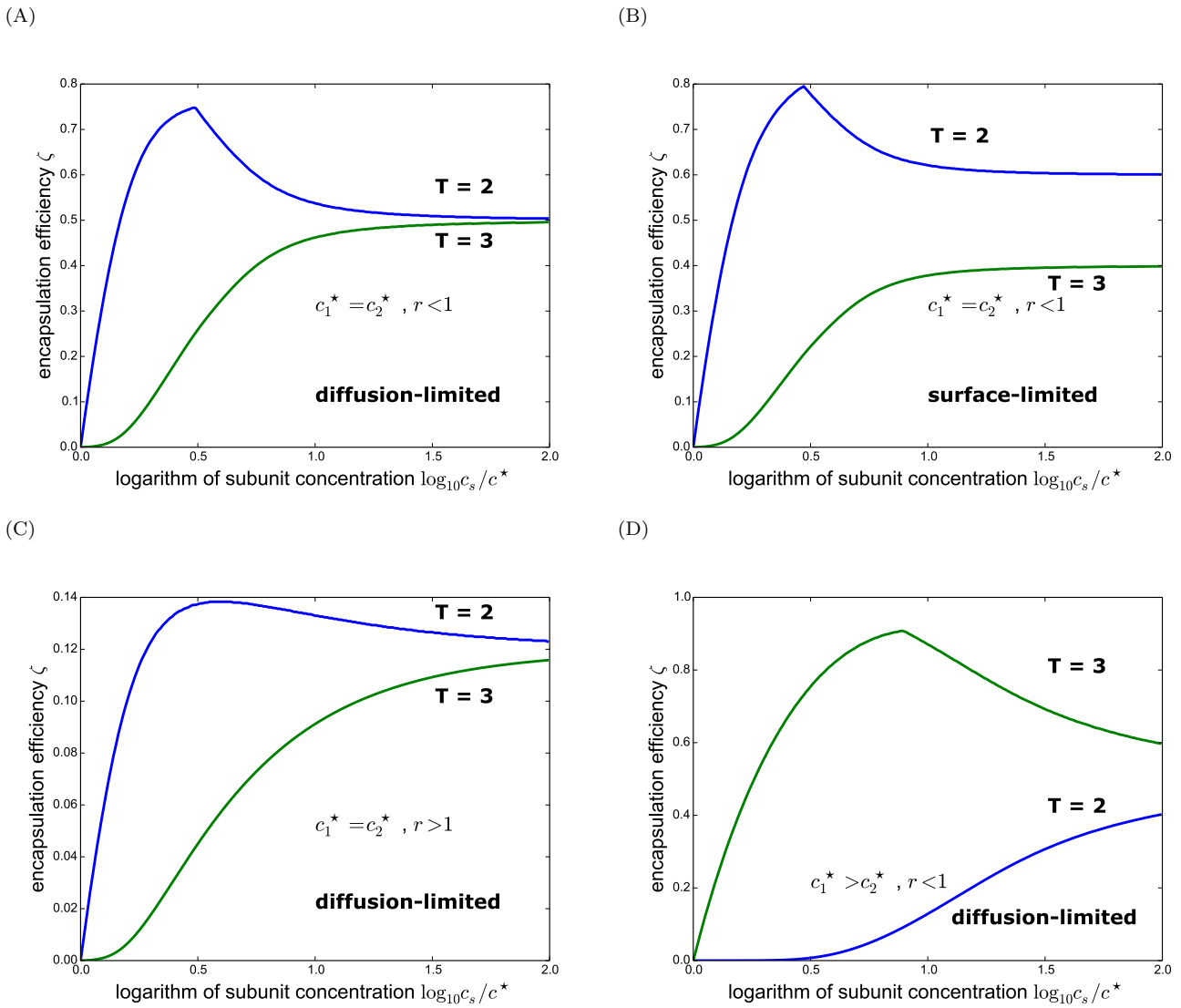


FIG. 10. Encapsulation efficiency ζ as a function of logarithm of subunit concentration $c_s/c^\star$, where $c^\star$ is the largest critical concentration, when two types of capsid T=2 and 3 can form and compete for protein and cores as predicted from kinetics. At shallow quenches assembly is monodisperse, while at deep quenches there is co-existence, which is the major result of our work, as can be seen in all figures. The level of co-existence is affected by the type of kinetics: diffusion-limited kinetics in figure (A) favor no capsid type, while surface-limited kinetics favor smaller types in figure (B). Additionally, the dependence on the stoichiometric ratio $\tilde{r} = q\, c_s/c_c$, is much weaker than the one predicted from thermodynamics, where c_c is the core concentration and q the number of subunits in a capsid. Figures (A) and (C) correspond to stoichiometric ratio r smaller and larger than 1 respectively. They are similar, apart from the sharp drop in efficiency of one type for $r < 1$, when efficiency of both types reaches the maximum value allowed so that they have to compete for protein. Smaller capsids are always preferred at shallow quenches, independent of stoichiometric ratio. In figure (D), where $c_1^\star > c_2^\star$, so that the larger type is energetically favored, the larger type dominates at shallow quenches, but at deep quenches co-existence is still preserved. The persistence of polymorphism, even when capsids have different critical concentrations, motivates the use of kinetics to describe capsid assembly.

state diagrams Fig.12A and Fig. 12B corresponding to equilibrium and kinetics, for equal critical concentrations $c_1^\star = c^\star$ and ratios $d^\star/c^\star$. The position of empty capsids on the diagrams follows from the case where only one filled capsid can form, but we assume the cases are similar apart from the redistribution of protein subunits between two types of filled capsid. We point out that this is a new type of state diagram for capsid assembly. We discuss the conclusions and outlook following our results in the last Sec. VII.

(A) (B)

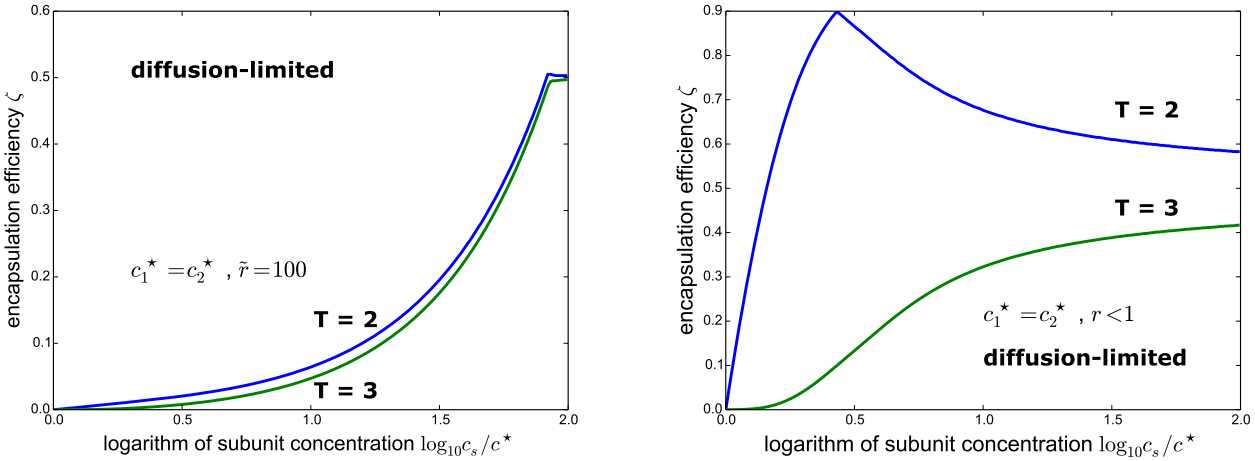


FIG. 11. Continued from Fig.10A-10D. In figure (A), the reduced core concentration $\tilde{r}_2 = q_2\, c_s/c_c = 100$ for T=3 is kept fixed. Co-existence occurs at a larger range of concentrations, since the stoichiometric ratio r is very large over most of it, so that the two types do not compete. In figure (B), the ratio $d^\star/c^\star$ is shifted, where $d^\star$ is the critical concentration for empty capsid assembly of the same T-number as the filled capsids corresponding to $c^\star$. Specifically, the ratio is made larger for $T = 2$, while neither type is made energetically favored. This corresponds to making the quench depth for the smaller capsids deeper, thus making assembly faster, as seen in Fig.6E. Co-existence is still observed at deep quenches. If the ratio keeps increasing, assembly can become mono-disperse, explaining why very large or very small T-numbers compared to the T-number of the wild-type virus are not observed in experiment.

(A) (B)

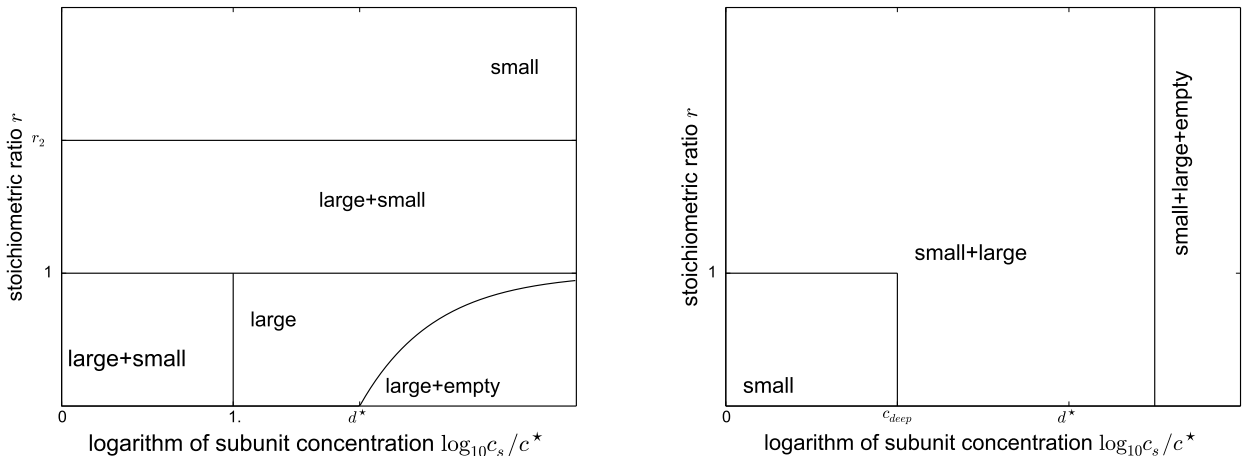


FIG. 12. Qualitative state diagrams of core-controlled assembly with parameters the logarithm of subunit concentration $\log_{10} c_s/c^\star$ and stoichiometric ratio $r = q\, c_s/c_c$, where q are the number of subunits in the largest appearing capsid and c_c the core concentration, as they are predicted from equilibrium theory in (A) and kinetic theory in (B). The different types are assumed to have the same critical concentration. Empty capsids assemble above a subunit concentration $d^\star$. The state diagram in Fig. **??** degenerates into Fig. 3A when one type becomes energetically favored. Equilibrium theory predicts a very narrow range of co-existence, that is for stoichiometric ratios above and below a stoichiometric ratio r_2 on the order of 1. Additionally, empty capsids appear only for stoichiometric ratios r below 1. However, the kinetic traps posed by empty and filled capsid of different T-number change the state diagram. Co-existence of T-numbers always occurs at deep quenches $c_s > c_{deep}$, where c_{deep} is subunit concentration $\sim 100 c^\star$ and empty capsids appear for stoichiometric ratios above 1, at a concentration c_s above the critical concentration for empty capsid assembly $d^\star$. The state diagrams motivate the idea, that in order to achieve a complete picture of polymorphism in capsid assembly, a larger range of concentrations $c_s/c^\star$ and stoichiometric ratio r have to be examined.

VII. DISCUSSION

Application of thermodynamics has been successful in describing features of core-controlled virus assembly[1, 2, 17, 33]. Specifically, it predicts a critical concentration above which capsids assemble, the high efficiency of assembly above that concentration as a function of concentration and stoichiometric ratio of proteins to cores as well as the selectivity of assembly [13, 37]. However, it assumes that protein subunits are able to distribute themselves between cores in order to achieve the state of minimum free energy in experimental time scales.

The free energy landscape of different T-numbers has been explored as function of core charge and radius [27]. The presence of kinetic traps in the assembly pathway could interfere with the system reaching the predicted minimum [42]. This may include malformed capsids [43] and as we suggest empty capsids and capsids of non-native T-number. In particular, if capsids of different T-number than the one that minimizes free energy assemble faster, then their disassembly could be extremely slow due to the hysteresis effect [44]. The particular pathway followed may also be prone to different kinetic traps.

The model by Zlotnick predicts hysteresis of capsid disassembly, an interesting property of virus capsids [45]. This is based on a stepwise-assembly pathway, where subunits are recruited in sequence to assemble the complete capsid. Hysteresis is also predicted by models based on the kinetic theory of phase transitions [46], which supports also the argument that capsid assembly is well described as a phase transition. Finally, coarse-grained Molecular Dynamics simulations by Rapaport [47] have shown that stepwise assembly with reversible binding does result into complete empty capsids, further motivating this model.

Core-controlled assembly may follow a different pathway, since the core acts as a nucleating center for the protein subunits, reducing the critical concentration required for assembly and allowing the subunits to first adsorb and then interact with each other in order to form a capsid [4]. This suggests an alternative pathway to capsid assembly, which we dub protein condensation pathway. Additionally, the type of cargo encapsulated has an impact on assembly. When RNA is used as the core-particle, assembly seems to undergo two steps. Firstly, a fast step where proteins adsorb to the genetic material, which then compactifies due to protein-protein interactions followed by recruitment at a slower rate of more protein that leads to completion of the capsid [48]. It is reported that assembly seems to be driven by specific interactions between the protein and its native RNA [49].

However, in an experiment by the UCLA virus group, CCMV RNA and BMV RNA had to compete for CCMV protein and the non-native RNA was reported to reach higher efficiencies than the native one [12]. Even though the two viruses are closely related, it seems that non-specific electrostatic interactions drive largely the assembly process. Still specific interactions seem to be important as well, enabling the assembly to follow a limited number of pathways that ensure fidelity of the T-number assembled [49]. The assembly mechanism of non-organic cores is similar to organic ones, apart from the compactification step, since the inorganic cores are already in an assembly-active rigid spherical form.

Kinetics, implemented here by using CNT and a circular-rim model, is able to predict such kinetic traps and possibly differentiate between different assembly pathways. Monte Carlo simulations of core-controlled assembly that allow for polymorphism have been done by Hagan [50], demonstrating the fact that T-number may vary with the size of the template and that kinetics compete with thermodynamics. Hagan also explores the results of a kinetic model following the example of Zlotnick for empty capsids [25]. This reproduces the critical charge density on the core-particle that is required for successful capsid assembly, high encapsulation efficiencies as well as the dependence on the stoichiometric ratio.

Our model indicates that equilibrium considerations and kinetics do differ in their predictions. Equilibrium theory predicts a strong dependence of polymophism on the stoichiometric ratio, while kinetics predicts a very weak one and a strong dependence on the quench depth $\Gamma \sim \log_{10} c_s/c^\star$. Capsids nucleate at the same time, and a complete capsid is assumed to be stable, so that the system is not allowed to reach the state of minimum free energy that equilibrium theory predicts. This contradiction is especially evident in the case when $r < 1$, where according to equilibrium theory the larger type is preferred so that as many subunits as possible benefit from being in a capsid, while kinetics allow for co-existence of different T-numbers.

Polymorphism, as predicted from kinetics, strongly depends on the quench depth $\Gamma \sim \log_{10} c_s/c^\star$. For shallow quenches, the barrier to nucleation depends strongly on T-number and assembly is monodisperse. According to the free energy difference between the two-types, either type dominates. For deep quenches, kinetics predicts co-existence of T-numbers, as observed in competition experiments [12, 21] as well as the appearance of empty capsids at unfavorable conditions [36]. Different assembly pathways, namely diffusion-limited and surface-limited that correspond to stepwise and protein-condensation assembly, dictate which T-number is expressed at larger quantity, possibly allowing to differentiate between assembly pathways. Measurement of time-lags may also help to determine the particular pathway followed.

The appearance of empty capsids and co-existence of different T-numbers in experiments of core-controlled capsid assembly implies that kinetics should be considered. However, it is possible that an alternative pathway of assembly

could counter-act this effect. Specifically, in an experiment by the Dragnea group [40], it was found that in the case that $r < 1$, subunits adsorb equally among cores, resulting into the frustration of assembly. The system escapes this trap by forming large complexes of cores and proteins, where the proteins are exchanged much faster between cores, resulting into complete capsids. This could potentially allow the system to reach its true equilibrium, although more experiments and modeling of the system are required to determine this.

Another argument against CNT kinetics is that, for deep quenches, the rate limiting step of assembly is not the critical nucleus, but the addition of the last subunit and closure of the rim. This has been shown in simulations for empty capsids [51]. Additionally, simulations of L-J disks [32] on a sphere have shown that the rim may close prematurely for small capsids, leading to a less stable aggregate than the complete capsid. This would change how kinetics behave at deep quenches, since complete capsids would appear at the rate rims are bridged. This has not been thoroughly examined, as the TEM micrographs usually produced in experiments do not provide enough resolution to discern if capsids are complete or not.

In conclusion, a complete picture of how capsid assembly depends on experimental parameters has not been drawn. The importance of the quench depth $\Gamma \sim \log_{10} c_s/c^\star$, and stoichiometric ratio r on the state diagram have been demonstrated. Determining the critical concentration is essential, in order to have a measure of the quench depth, as well as ranging the protein concentration over many orders of magnitude, because of the logarithmic dependence. While conditions, such as the size and charge of the core-template or acidity and salinity of the solution, have been thoroughly examined, we propose that they are not enough to have a reliable picture of capsid assembly. Systematic experiments covering a larger range of concentations and stoichiometric ratios and how this affects polymorphism would elucidate previous results and may indicate different mechanisms of capsid assembly. This in turn would lead to control over the products of assembly, which is of vital importance for applications.

ACKNOWLEDGEMENTS

We would like to thank Pepijn Moerman for the useful discussions.

[1] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, J. Rudnick, PNAS 101, 15556 (2004).
[2] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, R. Zandi, Phys. Rev. Lett. 90, 248101-1 (2003).
[3] P. Ceres, A. Zlotnick, Biochemistry-US 41,11525 (2002).
[4] A. McPherson, BioEssays 27, 447 (2007).
[5] L. Lavelle, M. Gingery, M. Phillips, W.M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, J. Ruiz-Garcia, J. Phys. Chem. B 113, 3813 (2009).
[6] M. Cuillel, C. Berthet-Colominas, P. A. Timmins, M. Zulauf, Eur. Biophys. J. 15, 169 (1987).
[7] R. W. Lucas, S. B. Larson, A. McPherson, J. Mol. Biol. 317, 95 (2002).
[8] A.C. Steven., J.F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J. Stahl, P. T. Wingfield, Adv. Virus Res. 64, 125 (2005).
[9] S. K. Dixit, N. L. Goicochea, M. C. Daniel, A. Murali, L. Bronstein, M. De, B. Stein, V. M. Rotello, C. C. Kao, B. Dragnea, Nano Lett. 6, 1993 (2006).
[10] C. Chen, M. C. Daniel, Z. T. Quinkert, M. De, B. Stein, V. D. Bowman, P. R. Chipman, V. M. Rotello, C. C. Kao, B. Dragnea, Nano Lett. 6, 611 (2006).
[11] R. D. Cadena-Nava, Y. Hu, R. F. Garmann, B. Ng, A. N. Zelikin, C. M. Knobler, W. M. Gelbart, J. Phys. Chem. B 115, 2386 (2011).
[12] M. Comas-Garcia, R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, W. M. Gelbart, J. Virol. 86, 12271 (2012).
[13] R. Zandi, P. van der Schoot, Biophys. J. 96, 2 (2009).
[14] Z. Varpness, P. A. Suci, D. Ensign, M. J. Young, T. Douglas, Chem. Commun. , 3726 (2009).
[15] F. Tseng, I. B. Tsvetkova, Y-Lan Khuong, A. W. Moore, R. J. Arnold, N. L. Goicochea, B. Dragnea, S. Mukhopadhyay, Mol. Pharmaceutics 10, 51 (2013).
[16] A. Escosura, R. J. M. Nolte, J. L. M. Cornelissen,J. Mater. Chem. 19, 2274 (2009).
[17] W. K. Kegel, P. van der Schoot, Biophys. J. 86, 3905 (2004).
[18] W. K. Kegel, P. van der Schoot, Biophys. J. 91, 1501 (2006).
[19] J. E. Johnsom, J. A. Speir, J. Mol. Biol. 269, 665 (1997).
[20] P. Csermely , R. Palotai, R. Nussinov, Trends Biomed. Sci. 35, 539 (2010).
[21] L. He, Z. Porterfield, P. van der Schoot, A. Zlotnick, B. Dragnea ACS NANO (2013).
[22] A. Zlotnick, N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, S. J. Stahl, P. T. Wingfield, Biochemistry-US 35, 7412-7421 (1996).
[23] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, D. Endres, Biochemistry-US 38, 14644 (1999).

[24] M. C. Daniel, I. B. Tsvetkova, Z. T. Quinkert, A. Murali, M. De, V. M. Rotello, C. C. Kao, B. Dragnea, ACS Nano 4, 3853, (2010).
[25] M. F. Hagan, J. Chem. Phys. 130, 114902 (2009).
[26] J. Sun, C. Dufort, M.C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, B. Dragnea, PNAS 104, 1354 (2007).
[27] A. Siber, R. Zandi, R. Podgornik, Phys. Rev. E 81, 051919 (2010).
[28] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. Duijn, A. J. R. Heck, Nat. Chem. 3, 126 (2011).
[29] M. F. Hagan, D. Chandler, Biophys. J. 91, 24 (2006).
[30] D. Kashchiev, Nucleation : basic theory with applications, BH (2000).
[31] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, H. Reiss, Biophys. J. 90, 1939 (2006).
[32] A. Luque, D. Reguera, A. Morozov, J. Rudnick, R. Bruinsma J. Chem. Phys. 136, 184507 (2012).
[33] A. Siber, A.Majdandzic, Phys. Rev. E 80, 021910 (2009).
[34] P. Debye, Ann. N.Y. Acad. Sci. 51, 575592 (1949).
[35] T. Keef, A. Taormina, R. Twarock, Phys. Biol. 2, 175 (2005).
[36] B. Dragnea, personal communication.
[37] I. Tsevtkova, C. Chen, S. Rana, C. C. Kao, V. M. Rotello, B. Dragnea, Soft Matter 8, 4571 (2012).
[38] D. T. Wu, Solid State Phys. 50, 37 (1997).
[39] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, D. T. Wu, Virology 325, 320327 (2004).
[40] A. G. Malyutin, B. Dragnea, Phys. Chem. B 117, 10730-10736 (2013).
[41] H. Ishikawa, K. Kwak, J. K. Chung, S. Kim, M. D. Fayer, PNAS 105, 8619 (2008).
[42] M. F. Hagan, O. M. Elrad, R. L. Jack, J. Chem. Phys. 135, 104115 (2011).
[43] Y. Ding, Y. P. Chuan, L. He, A. P. J. Middelberg, Biotechnol. Bioeng. 107, 550 (2010).
[44] C. Uetrecht, N. R. Watts, S. J. Stahl, P. T. Wingfield, A. C. Steven, A. J. R. Heck, PCCP 12, 13368 (2010).
[45] S. Singh, A. Zlotnick, Jour. Biol. Chem. 278, 1824918255 (2003).
[46] P. van der Schoot, R. Zandi, Phys. Biol. 4, 296304, (2004).
[47] D. C. Rapaport, Physics Procedia 34, 114 (2012).
[48] A. Borodavka, R. Tuma, P. G. Stockley, PNAS 109,15769-15774, (2012).
[49] E. C. Dykeman, P. G. Stockley, R. Twarock, J. Chem. Phys. 135, 104115 (2011).
[50] O. M. Elrad, M. F. Hagan Nano Lett. 8, 3850 (2008).
[51] H.D. Nguyen, V. S. Reddy, C. L. Brooks III, Nano Lett. 7, 338 (2007).