REVIEWS OF MODERN PHYSICS, VOLUME 93, APRIL–JUNE 2021

Equilibrium mechanisms of self-limiting assembly

Michael F. Hagan[*]
Martin Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02454, USA

Gregory M. Grason[†]
Department of Polymer Science and Engineering, University of Massachusetts,
Amherst, Massachusetts 01003, USA

(published 11 June 2021)

Self-assembly is a ubiquitous process in synthetic and biological systems, broadly defined as the spontaneous organization of multiple subunits (macromolecules, particles, etc.) into ordered multiunit structures. The vast majority of equilibrium assembly processes give rise to two states: one consisting of dispersed disassociated subunits and the other consisting of a bulk-condensed state of unlimited size. This review focuses on the more specialized class of *self-limiting assembly*, which describes equilibrium assembly processes resulting in finite-size structures. These systems pose a generic and basic question, how do thermodynamic processes involving noncovalent interactions between identical subunits "measure" and select the size of assembled structures? This review begins with an introduction to the basic statistical mechanical framework for assembly thermodynamics that is used to highlight the key physical ingredients ensuring that equilibrium assembly will terminate at finite dimensions. Then examples of self-limiting assembly systems are introduced, and they are classified within this framework based on two broad categories: *self-closing assemblies* and *open-boundary assemblies*. These include well-known cases in biology and synthetic soft matter (micellization of amphiphiles and shell and tubule formation of tapered subunits) as well as less widely known classes of assemblies, such as short-range attractive or long-range repulsive systems and geometrically frustrated assemblies. For each of these self-limiting mechanisms, the physical mechanisms that select equilibrium assembly size, as well as the potential limitations of finite-size selection, are described. Finally, alternative mechanisms for finite-size assemblies are discussed, and contrasts are drawn with the size control that these can achieve relative to self-limitation in equilibrium, single-species assemblies.

DOI: 10.1103/RevModPhys.93.025008

CONTENTS

---

[*]hagan@brandeis.edu
[†]grason@umass.edu

I. INTRODUCTION

A. Overview

Self-assembly is a process in which multiple subunits, or "building blocks," spontaneously organize into collective and coherent structures. This process is ubiquitous in living systems, where it underpins a wide range of structures at the cellular and subcellular scales, from lipid membranes to multiprotein filaments and capsules (Alberts *et al.*, 2002). Inspired by biology's successful strategies to build functional nanostructures, self-assembly is forming the basis of modern approaches to generate materials from the "bottom up" (Hamley, 2003). Chemical techniques enable the synthesis of a bewildering array of small-molecule, macromolecular, or particulate subunits that are engineered to self-assemble into high-order architectures (Klok and Lecommandoux, 2001; Stupp and Palmer, 2014; Boles, Engel, and Talapin, 2016). As in the biological context, the assemblies bridge between the scales of molecules and chemical function (nanometer and subnanometer) to size scales that are useful for controlling material properties (microns and beyond).

In different domains of science and engineering, the term "self-assembly" often connotes a range of distinct, if overlapping, physical processes. In its broadest usage, self-assembly implies the collective association of multiple elements into organized configurations, by dynamics that start from a relatively "disorganized" state and evolve with at least some degree of randomness. The great conceptual appeal of self-assembly in materials science is that the instructions for a desirable or useful structure may somehow be imprinted on the assembling subunits themselves, such that in a simple mixture the desired target structures emerge from the random processes of Brownian motion and subunit association.

In this review, we focus on *self-limiting assembly* (SLA), which is defined as self-assembly processes that terminate at an equilibrium state in which superstructures have a well-defined and finite spatial extent in one or more dimensions. Many examples of finite assemblies can be found in biological systems, where the assembly of identical subunits into larger, yet finite-sized, superstructures is common and functionally vital. As shown in Fig. 1, examples include (i) the protein shells that enclose viruses (Caspar and Klug, 1962; Mateu, 2013; Perlmutter and Hagan, 2015) and microcompartments (Tanaka *et al.*, 2008; Kerfeld, Heinhorst, and Cannon, 2010; Rae *et al.*, 2013), (ii) finite-size protein superstructures in photonic tissues (Prum *et al.*, 2009; Saranathan *et al.*, 2012; McPhedran and Parker, 2015), and (iii) finite-diameter bundles and fibers of cytoskeletal or extracellular protein filaments (Neville, 1993; Fratzl, 2003; Popp and Robinson, 2012). Each of these examples shares the notable feature that the finite size of the assembled structure far exceeds the nanometer size scale of the protein subunits. Crucial to their biological roles, the functional properties of these protein superstructures are regulated through the control of their finite size: respectively, (i) selective encapsulation and transport, (ii) optical response, and (iii) stiffness and strength. In this way, nature exploits self-assembly to deploy structures, built from the same or similar subunits, in diverse intracellular and extracellular environments and adapts their performance and functions by controlling the size of the assembled structure.

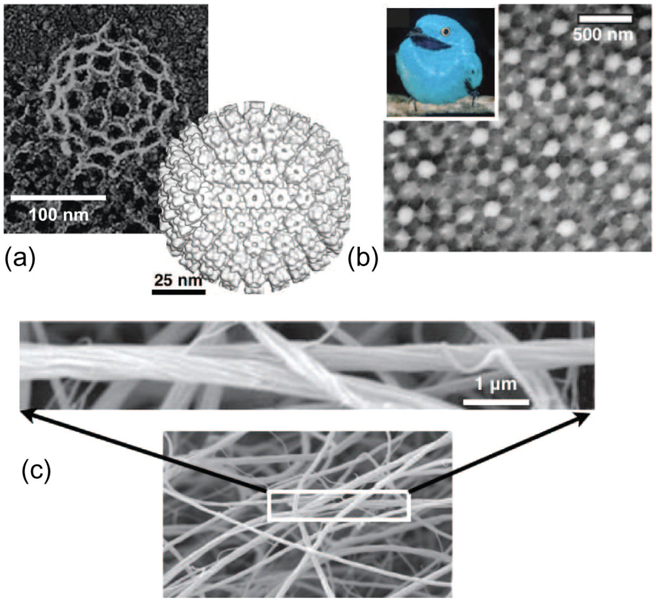

FIG. 1. Functional, finite-sized assemblies of proteins in biology: (a) protein shells of clathrin (left panel) and viral capsids of Herpes simplex (right panel), (b) photonic nanostructures formed by keratin aggregates in feather barbs of plum-throated continga (inset), and (c) finite-diameter fibers in reconstituted fibrin clot. (a), left panel: adapted from Royle, 2012. (a), right panel: adapted from Baker, Olson, and Fuller, 2000. (b): adapted from Dufresne *et al.*, 2009. (c): adapted from Weisel, 2004.

In contrast to these examples, most typical mechanisms of self-assembly in synthetic systems result in unlimited organized states, such as crystalline or liquid-crystalline mesophases. In these states, structure may be well defined on some microscopic scale, such as the unit cell dimension, but its overall size is uncontrolled by assembly thermodynamics. This result, which may be described as bulk phase separation, is a generic consequence of the thermodynamic trade-off between entropic and energetic drives. In the most general case, once the net cohesive drive for a subunit to join an assembled structure exceeds the entropic penalty for giving up its higher configurational freedom as a disassociated unit, there is no thermodynamic reason to stop this process. Thus, subunits continually add to the aggregate until it reaches macroscopic proportions and the subunits are nearly depleted.

This review describes the basic physical ingredients and common outcomes of assembly mechanisms that terminate at well-defined, finite sizes. We draw upon examples of SLA from biological systems, and consider the requirements to achieve such assemblies in synthetic systems. For clarity, we specifically focus on assemblies comprising a single species of identical subunits. Moreover, we restrict our definition of SLA to *equilibrium* assembly mechanisms, meaning that assembly terminates at a finite-sized *free energy* minimum structure.

Equilibrium assembly processes deserve special focus for both conceptual and practical reasons. A key advantage is that they are described by well-defined and generic statistical mechanical principles. This allows one, as we attempt to do in this article, to draw sharp distinctions between assemblies that either are or are not self-limiting. Reaching thermodynamic equilibrium requires subunits to associate and disassociate from aggregates sufficiently freely that a thermodynamically large collection of subunits behaves ergodically, sampling a sufficiently large ensemble of aggregation states in an experimentally relevant time. For systems at or near room temperature, such conditions are accessible when assembly is driven by noncovalent and reversible interactions, of the type that characterize physical association between macromolecules and colloidal particles in solutions (Russel, Saville, and Schowalter, 1989; Israelachvili, 2011), including van der Waals, electrostatic, hydrophobic, hydrogen-bonding, and depletion forces.

The ability of reversibly associating assemblies, if given sufficient time, to proceed toward one specific, thermodynamically defined state points to practical advantages of equilibrium assembly. As evidenced by the previously referenced synthetic approaches to size-controlled structures, nonequilibrium control over finite dimensions of assemblies requires extensive protocols to control the assembly environment, such as precisely regulating the temporal sequence of temperatures and subunit concentrations. This makes it exceedingly difficult, if not impossible, to deploy these nonequilibrium size-control strategies in uncontrolled environments, such as the complex and dynamic milieu of living organisms. In such scenarios where assembly cannot be carefully supervised, equilibrium mechanisms of assembly offer the distinct advantage that the final states may still be well defined. For example, viruses can exert only limited control over the intercellular media of their host organisms. Nevertheless, to be infectious, size-controlled capsid shells must assemble with high fidelity from the capsomer subunits. While this assembly process is in general not purely at equilibrium, biology often achieves such high fidelity by building upon equilibrium processes. For example, many viral capsids can spontaneously assemble from their purified components under (near) equilibrium conditions, with structures that are indistinguishable from capsids formed within a host cell (Wingfield *et al.*, 1995; Johnson and Speir, 1997; Fox *et al.*, 1998; Wang *et al.*, 2015), and in some cases are even infectious (Fraenkel-Conrat and Williams, 1955).

At the center of this specialized focus on equilibrium mechanisms for self-limiting single-species assembly is the following puzzle: how can equilibrium association processes “measure” the assembly to select a thermodynamically preferred state that is larger than a single subunit, yet less than infinite (i.e., bulk)? Because thermodynamic equilibrium is independent of the history of system, this state cannot be defined by the temporal process in which subunits arrive to the aggregate. Nor do these identical subunits have specific “addresses” that prescribe where they are supposed to sit in a particular aggregation state. The answers, not surprisingly, lie in how the shape and interactions of subunits conspire to determine the dependence of assembly energetics on size. For example, in the canonical example of SLA, formation of spherical micelles from amphiphilic molecules (Israelachvili, Mitchell, and Ninham, 1976), the assembly motif favors individual subunits to span the assembly from the solvophobic core to the solvophilic surface. Hence, in this case it is intuitive that energetics favors aggregates that are limited to sizes that are comparable to the lengths of the amphiphiles themselves. Far less intuitive is how single-species assemblies select finite equilibrium sizes that are much bigger than the subunit dimensions, or the range of their interactions. That is, what are limits of self-limitation, i.e., how large can a self-limited structure be, and how does this size limit depend on the physical characteristics (shape, interactions, etc.) of the subunits?

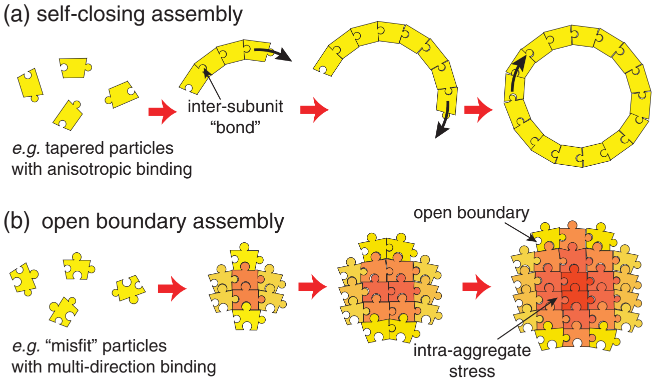

FIG. 2. Schematic illustrations of two classes of SLA described in Sec. III. (a) Self-closing assembly, in which inter-subunit rotations lead to cohesive assembly into closed, boundary-free aggregations. (b) Open-boundary (self-limiting) assembly, in which intra-aggregate stress accumulates with assembly and restrains the cohesive drive toward unlimited size.

B. Outline

With these basic questions in mind, this review has two broad aims. We first overview the generic statistical and thermodynamic elements of SLA, then present a classification for known mechanisms of SLA of identical subunits. The review is organized into two main sections based on these aims. In Sec. II, we present the thermodynamic principles of SLA based on the statistical mechanics of ideal aggregation of identical subunits. This begins with an introduction to ideal aggregation theory and illustration of the more generic case of *unlimited assembly*. Following this, we introduce a generic description of the ingredients for SLA. We review how the onset of aggregation, known as the critical aggregation concentration (CAC), the self-limiting size of aggregates, and the statistics of aggregate size fluctuations depend on the functional form for the size dependence of the intra-aggregate interaction energy. We then review the conditions for “polymorphic” SLA, in which assemblies exhibit multiple states of aggregration (some finite, some not). These systems are characterized by so-called secondary CACs, in which increasing concentration sufficiently far above the CAC leads to additional transitions between aggregation states.

Section III describes physical systems that exhibit SLA and classifies the models that capture their behavior into two categories illustrated schematically in Fig. 2: *self-closing* and *open-boundary* assembly. The former category describes assembly processes that terminate because they close upon themselves [Fig. 2(a)] and applies to shell and tubule formation, as well as the micellar assembly of surfactants, copolymers, and other amphiphiles. The last category applies to arguably lesser known classes of systems that have short-range attractions and long-range repulsions or form geometrically frustrated assemblies. These conditions enable assembly to terminate when the aggregate still has open boundaries characterized by a finite surface energy [Fig. 2(b)]. For this case, we introduce a generic framework for understanding how the interplay between intra-aggregate stress accumulation and aggregate surface energy controls the finite size of aggregates and the phase boundary between self-limiting and bulk aggregation.

While the core focus of this review is on the ingredients and outcomes of SLA from the point of view of thermodynamic equilibrium, the kinetic processes by which such systems reach equilibrium states (or, in some cases, do not) are essential to their study, particularly from the experimental point of view. A comprehensive review of kinetic limitations on assembly, which is relevant to both self-limited and unlimited assembly, is beyond the scope of this review. Nevertheless, we provide an introduction to some key considerations of assembly kinetics in Sec. IV. In that section, our purpose is to illustrate how those features of the assembly energetics that give rise to size selection in equilibrium influence the principle kinetic pathways of their formation.

Before concluding, we provide a discussion in Sec. V of physical mechanisms leading to finite-size aggregates that fall outside the primary scope of the review, namely, nonequilibrium and multispecies SLA, and questions these pose to the forgoing discussion for the more limited focus on single-species equilibrium SLA. We conclude with some remarks about open challenges in the application of the mechanisms and principles of SLA.

C. Scope of review

This review considers equilibrium assembly mechanisms that terminate at well-defined, finite sizes. As this focus suggests, we leave out discussion of nonequilibrium processes in general and, more specifically, what might be called active-assembly processes, such as the steady-state length of treadmilling and severing cytoskeleltal filaments (Desai and Mitchison, 1997; Mohapatra *et al.*, 2016; Pollard, 2016). Beyond that, we specifically consider assembly mechanisms of a single species of identical subunits. To be sure, this leaves out an emerging and interesting area of research on so-called addressable assemblies (Jacobs and Frenkel, 2016; Zeravcic, Manoharan, and Brenner, 2017), where mixtures of multiple distinct subunit species may be "programmed" to assemble into a specifically defined 3D structure in equilibrium. In this review, we provide only a limited discussion of size-controlled multispecies assembly and possible trade-offs with single-species mechanisms, particularly how the number of required species increases with target size.

Although the fabrication and synthesis of finite, size-controlled structures is well known in synthetic materials, such as size-controlled nanoparticles of atoms (Yin and Alivisatos, 2005; Cozzoli, Pellegrino, and Manna, 2006) and macromolecules (Hiemenz and Lodge, 2007), these examples raise a key distinction between equilibrium and nonequilibrium assembly. The control over finite size in all of these foregoing examples relies on the nonequilibrium process by which they form. For example, the size distribution of metal nanoparticles (Yin and Alivisatos, 2005; O'Brien, Jones, and Mirkin, 2016) is selected through spatiotemporal control of the physical-chemical factors that control nanocrystal growth (concentrations, temperature, ionic conditions, etc.). Indeed, as we later discuss, in generic conditions under which such assemblies form, allowing these assemblies to proceed to thermodynamic equilibrium would destroy the size control. Finite sizes are possible only when these processes are driven, maintained, and arrested out of equilibrium. In this sense, we reserve the term "self-limiting" for those rarefied assembly processes that result in finite-size structures in thermodynamic equilibrium. The physical mechanisms of equilibrium assembly that achieve such size control are the central focus of this review.

II. THERMODYNAMIC ELEMENTS

We begin with a review of the elementary statistical mechanical framework to describe equilibrium aggregation. We then illustrate the statistical thermodynamics of aggregation in models of what we will call *canonical aggregation*, where assembly proceeds via cohesive short-range and stress-free assembly of elemental units into 1D, 2D, and 3D aggregates. We illustrate how so-defined canonical assemblies do not exhibit self-limitation. We then describe the generic conditions for self-limiting (finite) equilibrium assembly and give an overview of the concentration-dependent thermodynamics of self-limiting assembly. Finally, we discuss models of competing finite aggregates and polymorphic transitions between finite and unlimited assembly, both of which may be characterized by multiple aggregation thresholds in the ideal theory.

A. Equilibrium principles

In this review we concern ourselves with equilibrium association of single subunits, or *monomers*, into states with aggregation number n subunits, or n-mers. Our purpose is to describe the minimal ingredients of assembly dominated by structures with finite aggregation number n. To this end, we restrict our presentation to ideal aggregation theory, where interactions between distinct aggregates are neglected. This is not to say that interactions among subunits within the same aggregate are neglected; quite the contrary. As we describe later, the intra-aggregate energetics and its n-dependence are critical for determining whether association leads to self-limited states or, instead, to more canonical states of bulk aggregation.

1. Classical aggregation theory: Fixed total concentration, noninteracting aggregates

Ideal aggregation theory is well established for certain classes of self-assembly systems, particularly in the context of amphiphiles and surfactants. As such, this theory was better described, in greater depth, by Tanford (1974), Israelachvili, Mitchell, and Ninham (1976), Gelbart, Ben-Shaul, and Roux (1994), and Safran (1994). Here our purpose is to consider the application and implications of ideal aggregation theory to a broader class of self-limiting assembly systems. Hence, we present only a minimal introduction to the elements necessary to describe aggregation to self-limiting states.

We consider a solution of N total subunits in a fixed total volume V. In what follows, we refer to unassembled single subunits as monomers.[1] To describe the concentration of subunits, it is convenient to scale concentration by the reference state concentration v_0^{-1}; i.e., v_0 is the volume per subunit in the reference state.[2] In this way, we can define concentration in nondimensional terms as the total volume fraction of subunits, $\Phi = N v_0/V$.[3] The N subunits are distributed among distinct n-mer aggregates, with the volume fraction of subunits in n-mers defined as ϕ_n. In the following, we refer to ϕ_n as the subunit fraction distribution. Throughout this review, it is also useful to introduce a separate variable for the aggregate distribution $\rho_n \equiv \phi_n/n$, which describes the relative count of n-mers in the mixture. Defined in this way, ϕ_n and ρ_n are less than unity for all n. Moreover, for the particular assumptions of ideal aggregation to hold (i.e., two-body contacts between aggregates are vanishingly rare), these quantities must all remain much less than unity. To simplify the nomenclature, throughout the review we refer to the nondimensional concentration (volume fraction) simply as the concentration.

We define $n\epsilon(n)$ as the free energy of intra-aggregate interactions; i.e., $\epsilon(n)$ is the per subunit aggregation free energy in an n-mer. The total free energy F for the ideal distribution of aggregates is given by

$$\frac{F}{(V/v_0)} = \sum_{n=1}^{\infty} \phi_n \left(\epsilon(n) + \frac{k_B T}{n} [\ln(\phi_n/n) - 1] \right), \tag{1}$$

with the two terms in the parentheses, respectively, representing the intra-aggregate interaction free energy and translational entropy (in the ideal solution approximation) of n-mers, with the $1/n$ in the latter term reflecting the critical fact that all subunits of an n-mer share a common, single center-of-mass degree of freedom.

To obtain the equilibrium aggregate size distribution, we minimize F with respect to ϕ_n subject to the constraint that the total subunit concentration is fixed:

$$\sum_{n=1}^{\infty} \phi_n = \Phi, \tag{2}$$

i.e.,

$$\frac{\partial}{\partial \phi_n} \left[F + \mu \left(\Phi - \sum_{n=1}^{\infty} \phi_n \right) \right] = 0, \tag{3}$$

with μ playing the role of a Lagrange multiplier. This yields

$$\begin{aligned} \mu &= \epsilon(1) + k_B T \ln \phi_1 = \epsilon(2) + \frac{k_B T}{2} \ln(\phi_2/2) \\ &= \ldots = \epsilon(n) + \frac{k_B T}{n} \ln(\phi_n/n) = \frac{\partial F}{\partial \phi_n}, \end{aligned} \tag{4}$$

showing that μ is the subunit chemical potential. This condition requires that subunits have the same chemical potential in all aggregates, and it derives from both the energetics of the assembly and the ideal translational entropy of the n-mer. Note that the prefactor of $1/n$ of the translational entropy, derived from the sharing of a single center of mass in an n-mer, reflects a generically higher translational entropy of disaggregated states. The limit $n \to \infty$ gives $\mu = \epsilon(\infty)$, describing the equilibrium between monomers and a bulk phase-separated condensate that has no translational entropy. For simplicity, throughout this review we choose to define energies such that $\epsilon(1) = 0$, in which case $\epsilon(n)$ is defined as the difference of the per subunit energy between an n-mer and disassembled monomers.

It is convenient to use the first equality in Eq. (4) to recast the chemical potential in terms of the unknown monomer concentration, from which we can reformulate the generic chemical equilibrium conditions in terms of the *law of mass action*

$$\phi_n = n(\phi_1 e^{-\beta\epsilon(n)})^n, \tag{5}$$

where $\beta^{-1} \equiv k_B T$. Inserting the expression for $\phi_n(\phi_1)$ into the fixed number concentration [Eq. (2)] and summing over n then results in an equation of state relating the total concentration Φ to the monomer concentration ϕ_1. This equation of state and the underlying distribution of assemblies ϕ_n derive from the specific n dependence of aggregate interactions, with equilibrium states that are dominated by self-limited aggregates occurring only for certain forms of $\epsilon(n)$.

2. Unlimited assembly: Short-range cohesive aggregation

Before describing models that give rise to SLA, we consider the thermodynamics of the simplest models of physical association, described by short-range attractions between subunits. While these models are relevant to a broad range of physical scenarios like colloidal crystallization (Manoharan, 2015; Morphew *et al.*, 2018), they do not exhibit SLA. Yet they serve as a useful reference point for illuminating the necessary conditions for SLA. Specifically, these models result in either a single dispersed state whose most populous aggregate state is $n = 1$ (the free monomer) or coexistence between the dispersed monomer-dominated state and an

---

[1] In literature on amphiphile aggregation the term “unimer” is often used to describe the single subunit to avoid overlap with the connotation of “monomer” as the chemical repeat of macromolecular chain, which is often a component of self-assembling molecular subunits.

[2] For concreteness, we may take v_0 to be the subunit volume in the disassociated state, which provides a convenient and nondimensional measure of concentration that will be much less than unity for dilute conditions. However, the formalism is independent of choice of reference state; for example, $v_0 = 1/N_{\mathrm{AV}}$ L with the reference state concentration of 1 mol/L commonly used in the life sciences. Changing the definition of the reference state just uniformly rescales the free energy of intra-aggregate interactions $\epsilon(n)$.

[3] Throughout, we also refer to Φ more colloquially as the “concentration” of subunits.

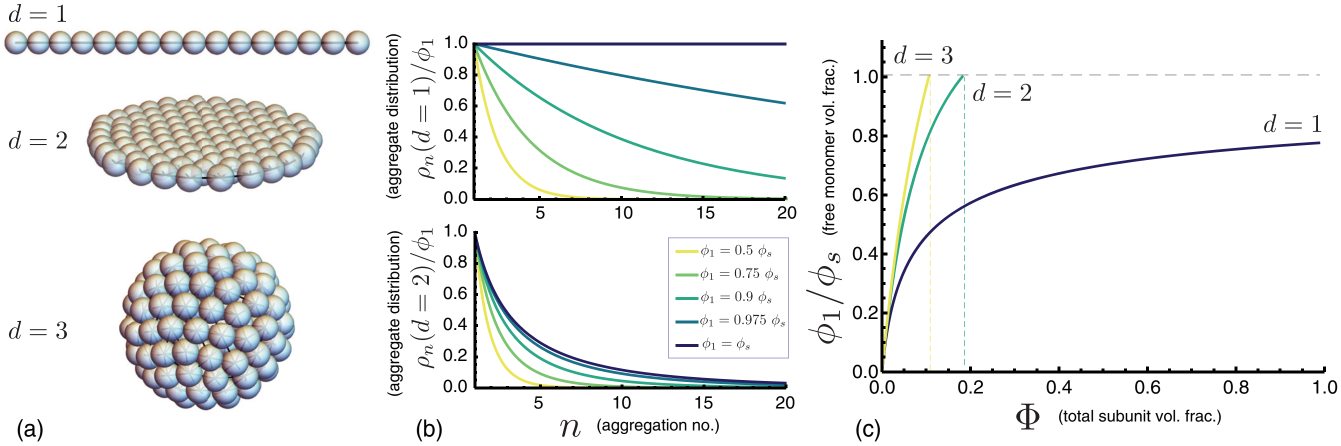

FIG. 3. (a) Examples of d-dimensional (linear, planar, and spherical) short-range cohesive aggregation. (b) Plots of the aggregation distributions (relative counts of n-mers) for $d = 1$ (top panel) and $d = 2$ (bottom panel) for monomer concentrations increasing to saturation (i.e., $\phi_1 = \phi_s$) for $\Delta_0 = 1k_BT$. While the dispersity (and mean size) of linear aggregates diverges as $\phi_1 \to \phi_s$, it remains finite for $d \geq 2$ at saturation. (b) Equation of state (ideal aggregation theory) for the free monomer population ϕ_1 as a function of the total concentration Φ for linear, planar, and spherical aggregates for $\Delta_0 = 2.75k_BT$. For $d = 2$ and $d = 3$ the free monomer concentration saturates at a finite Φ where $\phi_1 = \phi_s$. For linear aggregation, saturation is not reached in the ideal theory.

unlimited aggregate (macrophase separation). The absence of equilibrium finite-sized states will be traced to the generic size dependence of the cost of open boundaries at the edges of cohesive aggregates. In subsequent sections we show mechanisms that compete against the open-boundary cost to enable SLA.

Here we consider models where inter-subunit association promotes uniform d-dimensional aggregates, such as one-dimensional chainlike or two-dimension sheetlike aggregates. Every internal subunit forms, on average, z attractive bonds of strength $-u_0$, and subunits at the free boundary have δz fewer contacts [Fig. 3(a)]. For example, in a $d = 1$ chainlike assembly $z = 2$ and $\delta z = 1$. For generic dimensionality, the interaction free energy takes the form

$$\epsilon(n) = -\epsilon_0 + \frac{\Delta_0}{n^{1/d}}, \tag{6}$$

where $\epsilon_0 = u_0 z/2$ and represents the per subunit cohesive free energy in the bulk (i.e., $n \to \infty$) structure. The second term derives from the growth in the number of particles at the boundary ($\sim n^{(d-1)/d}$) and their deficit of cohesive bonds (δz), so that Δ_0 is equal to $u_0(\delta z)$ times a geometric factor accounting for the mean bond geometry at the boundary. The bonding geometry in these assemblies permits the structure to grow uniformly without disrupting this local contact structure at any size scale, a condition that we revisit when describing examples of self-limiting assembly in Sec. III.

We define the concentration $\phi_s \equiv e^{-\beta\epsilon_0}$, so that the law of mass action, Eq. (5), takes the form

$$\phi_n = n(\phi_1/\phi_s)^n e^{-\beta\Delta_0 n^\alpha}, \tag{7}$$

where $\alpha = 1 - 1/d$ is an exponent that characterizes the geometric growth of the exposed boundary with n. As shown in Fig. 3(b), $\phi_1 \leq \phi_s$ and the distribution $\rho_n = \phi_n/n$ decreases exponentially with aggregate size for large n, and for any d. Hence, under these conditions Φ, the sum over the subunit fraction distribution in Eq. (2), is finite, implying the existence of conditions where the concentration of subunits achieves equilibrium in the suspension. However, no such equilibrium exists for $\phi_1 > \phi_s$, implying that $\phi_1 \to \phi_s$ is an upper limit to concentrations that may be in equilibrium in a dispersed state. In other words, when Φ is sufficiently large that $\phi_1 = \phi_s$, the solution is *saturated*, and additional subunits (further increasing Φ) must phase separate to the macroscopic state (i.e., $n \to \infty$).

First consider the linear case ($d = 1$), where the equation of state can be readily computed from Eq. (7) with $\alpha = 0$ and the geometric series

$$\Phi = e^{-\beta\Delta_0} \frac{\phi_1 \phi_s}{(\phi_s - \phi_1)^2} \quad \text{for } d = 1, \tag{8}$$

which diverges as $\phi_1 \to \phi_s$. As plotted in Fig. 3(c), this divergence indicates that the monomer concentration increases with total concentration but never reaches the point of saturation (i.e., $\phi_1 < \phi_s$ for any finite Φ). Hence, for all subunit concentrations the system maintains $\phi_1/\phi_s < 1$, implying that the distribution of linear aggregates is always exponential, $\phi_n/n \propto e^{-n/\langle n \rangle}$, where the number-average length is $\langle n \rangle = 1/\ln(\phi_s/\phi_1)$.[4] Noting that $\epsilon_0 = \Delta_0$ for a 1D chain assembly, the growth of mean length with end energy in the limit of high concentration $\langle n \rangle \simeq e^{\beta\Delta_0/2}\sqrt{\Phi}$, which is well known for equilibrium polymers (Hiemenz and Lodge, 2007) and cylindrical micelles (Gelbart, Ben-Shaul, and Roux, 1994; Safran, 1994), highlights the mechanism that prevents "bulk" assembly for 1D aggregation. In this dimension, the probability of introducing a free end remains finite $\sim e^{\beta\Delta_0}$ in the $n \to \infty$ limit, analogous to the statistics of

---

[4] The *number-average* aggregate size is $\langle n \rangle = (\sum_n n\rho_n)/(\sum_n \rho_n) = (\sum_n \phi_n)/(\sum_n \phi_n/n)$, where ϕ_n/n is the distribution of n-mers, while the *mass-average* aggregate size is $\langle n \rangle_{\rm M} = (\sum_n n\phi_n)/(\sum_n \phi_n)$.

domain walls in the 1D Ising model at finite temperature (Fisher, 1984). However, while the mean size is finite, this case is distinct from what we describe here as self-limiting assembly, in both the strong dependence of $\langle n \rangle$ on total concentration and, perhaps more significantly, the fact that fluctuations in aggregate size are always comparable to the mean; that is, $\langle (n - \langle n \rangle)^2 \rangle^{1/2} \propto \langle n \rangle$.

For higher assembly dimensionality ($d > 1$), the geometric growth of the free boundary cost restrains the $n \to \infty$ divergence in the distribution [Eq. (7)] as the solution approaches saturation. For $\phi_1 = \phi_s$ the distribution takes the form $\phi_n(\phi_1 \to \phi_s) = n e^{-\beta \Delta_0 n^{\alpha}}$, the sum over which converges for $\alpha > 0$ when $d > 1$. For example, we can approximate the sum for planar aggregates ($d = 2$) by replacing the sum over aggregation number n with dimensionless aggregate radius $n = \pi r^2$, i.e., $\sum_n \phi_n \to 2\pi \int dr\, r\, \phi(r)$. At saturation, aggregate sizes are exponentially distributed, $\phi(r) \simeq \pi r^2 e^{-\sqrt{\pi}\beta\Delta_0 r}$, yielding a saturation concentration

$$\Phi_s(d = 2) \simeq 12 \left(\frac{k_B T}{\Delta_0} \right)^4, \qquad (9)$$

at which point the ideal solution of aggregates reaches equilibrium with the bulk condensate. Thus, with the exception of the special case of the $d = 1$ assembly, short-range cohesive aggregation is characterized by a finite saturation concentration $\Phi_s = \Phi(\phi_1 = \phi_s)$, above which subunits phase separate into an unlimited bulk structure.

The thermodynamics of these examples are plotted in Fig. 3(c) in terms of the equation of state $\phi_1(\Phi)$ for each dimensionality. Note that while the distributions of short-range interacting systems have finite mean sizes in the absence of macrophase separation, their distributions are dominated by monomers; i.e., the concentration of n-mers ρ_n is always maximal for $n = 1$, a property that sharply contrasts with the SLA behavior described next in Sec. II.B.

B. Self-limiting assembly: Elements and outcomes

Here we describe the generic ingredients and thermodynamic outcomes of assembly models that exhibit *self-limitation*. That is, unlike the previously described short-range cohesive models, these systems undergo ideal assembly into self-limiting states dominated by aggregates with a finite size n_* that is larger than 1, yet smaller than bulk unlimited states.

The physical mechanisms that give rise to this behavior are discussed in detail in Sec. III. Here we give an overview of the essential thermodynamic ingredients and behavior based on a generic description of the energetics of a self-limiting system. We consider the assembly behavior in terms of a generic function for the interaction free energy per subunit $\epsilon(n)$, which as sketched in Fig. 4(a) favors aggregation at a particular finite size, or possibly several distinct finite sizes. To highlight its distinct role from the translational entropy of aggregation, we use the term *aggregation energetics* to refer to $\epsilon(n)$, but we note that this describes a free energy per subunit, as interactions in general have both energetic and entropic contributions.

Given a known form of the $\epsilon(n)$, the discussion in this section addresses several key questions. First, what determines the onset of aggregation from the dispersed state? Second, what selects the dominant size of finite aggregates? And third, what are the conditions for driving transitions between different states of self-limiting aggregation, or between self-limited and unlimited aggregation states?

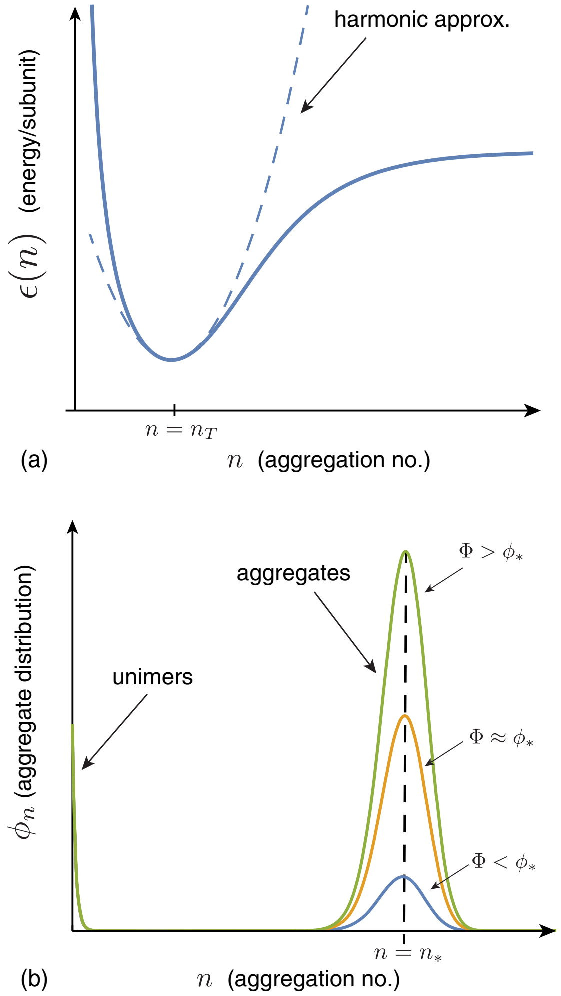

FIG. 4. (a) Schematic plot of the aggregation free energy per subunit shown as a continuous function of aggregation number n. The dashed line shows a harmonic expansion around a local minimum at the target size $n = n_T$. (b) Schematic plots of the aggregate distribution for a model of the type sketched in (a).

1. Aggregation threshold

We first describe the simplest picture of the concentration dependence of ideal aggregation. We begin with the assumption of an energy per subunit $\epsilon(n)$ of the form shown in Fig. 4(a), which has a single energy minimum at a finite aggregation number n_T, which we call the target size. The basic dependence of aggregation on concentration for such a model is sketched in Fig. 4(b). There are two dominant populations of aggregates, monomers and n-mers, with the n-mers narrowly distributed around the most populous state $n_* \approx n_T$.

For large enough n_*, the thermodynamics of aggregation can be captured, to a first approximation, by a *two-state*, or bimodal, distribution in which fluctuations around free monomers and the n-mer aggregate peak are neglected. We self-consistently test the validity of this approximation later; see Eq. (18). When subunits are distributed strictly between the $n = 1$ and $n = n_*$ states, the conservation of subunit mass is simply $\Phi = \phi_1 + \phi_{n_*}$. Chemical equilibrium then gives the concentration in preferred aggregates $\phi_{n_*} = n_*(\phi_1 e^{-\beta\epsilon_*})^{n_*}$, where $\epsilon_* \equiv \epsilon(n_*) < 0$ is the per subunit energy gain upon aggregation into the optimal size. Defining the concentration scale

$$\phi_* \equiv [n_* e^{-n_*\beta\epsilon_*}]^{-1/(n_*-1)} \approx e^{\beta\epsilon_*}/n_*^{1/n_*} \tag{10}$$

yields the following equation of state, relating total concentration to monomer concentration:

$$\frac{\Phi}{\phi_*} = \frac{\phi_1}{\phi_*} + \left(\frac{\phi_1}{\phi_*}\right)^{n_*}, \tag{11}$$

a result derived originally by Debye to explain scattering in soap solutions (Debye, 1949). The dependence of the populations of monomers and n_*-mers on total concentration is plotted in Fig. 5(a) and can be summarized as follows. For low concentration, $\Phi \ll \phi_*$ and additional subunits added to the system go predominantly to monomers since $\phi_1 \approx \Phi$, while $\phi_{n_*} \approx \phi_*(\Phi/\phi_*)^{n_*} \ll \Phi$. The population of aggregates in this regime ϕ_{n_*} is simply proportional to the random probability of n_* free subunits spatially coinciding Φ^{n_*} times the enhanced Boltzmann factor for aggregation $e^{-n_*\beta\epsilon_*} \approx \phi_*^{-n_*}$, and hence it is diminishingly small. In the large concentration regime $\Phi \gg \phi_*$, the dominant populations are reversed: the n_*-mer population increases in proportion to total concentration ($\phi_{n_*} \approx \Phi$), while monomers increase much more slowly [$\phi_1 \approx \phi_*(\Phi/\phi_*)^{1/n_*} \ll \Phi$]. These two regimes are characterized by a crossover near $\Phi \approx \phi_*$, which is known as the CAC, although it is not strictly a phase transition for finite n_*.[5] As illustrated in Fig. 5(a), the aggregation crossover becomes increasingly sharp as the aggregation number n_* increases. Figure 5(b) shows an example of CAC behavior observed in experiments with hepatitis B virus capsid protein assembly (Ruan, Hadden, and Zlotnick, 2018).

Underlying the transition is a thermodynamic trade-off between translational entropy and the interaction free energy that drives aggregation. Maximizing the ideal translational entropy of aggregates favors maximization of the number of independent translational degrees of freedom, i.e., the number of independent centers of mass in the mixture. To form an aggregate, n_* monomers must give up their n_* centers of mass for the single center of mass of the aggregate. Only when the aggregation free energy is sufficient to “pay” this entropic price (i.e., when the concentration of “excess” monomers is sufficiently large) does aggregation become thermodynamically favorable. Hence, the CAC depends on not only aggregation energetics but also the aggregation number n_*. According to Eq. (11), ϕ_* exhibits a modest increase with n_* (as $\sim n_*^{-1/n_*}$) due to the increased translational entropy loss for when joining larger aggregates. We return to the implications of the n_* dependence of aggregation thresholds in the discussion of competing aggregate states later on. Notice that, although the model ignores physical interactions between distinct aggregates, the change in translational entropy couples n_* units, making aggregation a cooperative process. For this reason, the CAC becomes progressively sharper and tends toward a thermodynamic transition as $n_* \to \infty$ [Fig. 5(a)].

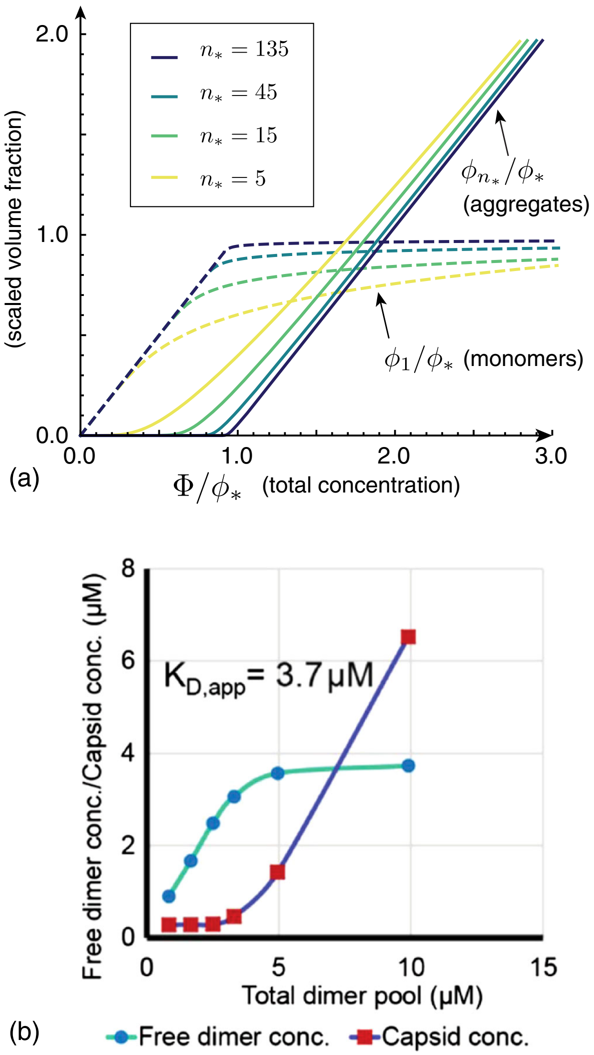

FIG. 5. (a) Plots of a “two-state” model composed only of monomers ($n = 1$) and aggregates of a single peak size ($n = n_* \approx n_T$) as functions of the total concentration Φ and for different finite aggregate sizes. The *critical aggregation concentration* (CAC), here ϕ_*, characterizes the concentration range beyond which aggregates dominate the subunit population. (b) Assembly behavior of hepatitis B virus (HBV) capsid protein *in vitro* as a function of concentration. The concentration dependence of the fraction of subunits (protein dimer) is shown in two states: free subunits and assembled capsids composed of 120 subunits. Notice that as the total concentration crosses the CAC (labeled $K_{D,\mathrm{app}}$) the concentration of free subunits is nearly constant, with almost all additional subunits assembling into capsids. From Ruan, Hadden, and Zlotnick, 2018.

---

[5] The CAC is commonly referred to as the *critical micelle concentration* (CMC) in the amphiphile literature. In the virus assembly literature it is often called the *pseudocritical concentration* to emphasize that it does not correspond to a true phase transition for finite n_*, and because the CAC observed in finite-time experiments typically exceeds the equilibrium CAC due to nucleation barriers.

2. Finite aggregates: Mean size and size dispersity

Here we review the conditions for the most probable or optimal aggregate size n_* given a known form of $\epsilon(n)$, which we assume for the moment to have a single minimum at target size $n_{\rm T}$. The optimal size n_* corresponds to the maximum in the aggregate distribution $\rho_n = (e^{\beta[\mu-\epsilon(n)]})^n$ or, equivalently, the minimum in the free energy $n[\epsilon(n)-\mu]$, which includes the total interaction free energy and entropy cost of forming an n-mer from free monomers. However, except under conditions where monomers are buffered to a fixed concentration, the chemical potential $\mu = k_B T \ln \phi_1$ varies as the equilibrium monomer population changes with total concentration. Naively, this might suggest that the optimal aggregate size should strongly vary with total concentration. Here we illustrate why, notwithstanding the variation of μ with concentration, n_* is nearly independent of Φ and almost entirely determined by the form of aggregation energetics $\epsilon(n)$. Following this, we summarize the effects of dispersity [i.e., the finite width of the aggregation peaks in Fig. 4(b)] on aggregation thermodynamics, which is necessary to account for the weak concentration dependence of the optimal aggregate size.

a. Two-state aggregation

As a first approximation, consider the two-state aggregation model deep into the aggregation regime, i.e., well above the CAC ($\Phi \gg \phi^*$). The optimal aggregate size derives from the condition $d\rho_n/dn|_{n_*} = 0$, or

$$\epsilon(n_*) + n_* \epsilon'(n_*) - \mu = 0, \tag{12}$$

where $\epsilon' = d\epsilon/dn$. Using the fact that $\phi_1(\Phi \gg \phi_*) \simeq \phi_*(\Phi/\phi_*)^{1/n_*} = e^{\epsilon_*}(\Phi/n_*)^{1/n_*}$ from Eq. (11) in Sec. II.B.1, this transforms to the condition for the optimal (peak) aggregate size

$$\epsilon'(n_*) = \frac{k_B T}{n_*^2} \ln(\Phi/n_*) \quad \text{(two state)}. \tag{13}$$

From Eq. (13) we may draw two key conclusions. First, in the limit of large target size $n_* \gg 1$, the optimal size corresponds to a minimum of $\epsilon(n)$. That is, since $\epsilon'(n_*) \to 0$, $n_* \to n_{\rm T}$ and the aggregate peak is selected by minimizing the subunit aggregation energy, independent (to a first approximation) of concentration. Second, the right-hand side of Eq. (13), which is proportional to the translational free energy of a dilute concentration of n_*-mers, is negative, and hence $\epsilon'(n_*) < 0$. Combining this with Eq. (12), we find

$$\mu < \epsilon(n_*). \tag{14}$$

Equation (14) shows that the equilibrium chemical potential approaches from below, but never quite reaches, the interaction energy of the optimal aggregate $\epsilon(n_*)$ (excepting the unphysical limit $\Phi/n_* \to 1$).

Finally, the fact that n_* corresponds to a *maximum* in the aggregate size distribution suggests the following condition from $d^2\rho_n/dn^2 < 0$:

$$\epsilon''(n_*) > -\frac{2\epsilon'(n_*)}{n_*} > 0. \tag{15}$$

As the right-hand side goes to zero as $\sim n_*^{-3}$, the aggregation energetics must be convex in the vicinity of the optimal size. Strictly speaking, however, a stronger condition than convexity alone is needed to justify the neglect of aggregation number fluctuations in the two-state approximation, as discussed next.

b. Gaussian approximation

We now consider the effect of *convexity* of the aggregation energetics, characterized by the second derivative of $\epsilon(n)$ at the peak aggregate size. As before, we restrict our analysis to the case of a single, well-defined minimum in $\epsilon(n)$ occurring at a finite target size $n_{\rm T} > 1$. Close to the minimal-energy size, the energetics have the form

$$\epsilon(n) \simeq \epsilon_{\rm T} + \frac{\epsilon''_{\rm T}}{2}(n - n_{\rm T})^2, \tag{16}$$

where $\epsilon_{\rm T} < 0$ and $\epsilon''_{\rm T} > 0$, respectively, characterize the minimum energy and convexity of the target aggregate, as illustrated in the harmonic approximation in Fig. 4(a). Physically, $\epsilon''_{\rm T}$, which we call the convexity, quantifies twice the energetic cost to alter the aggregate number from its target by ±1. In Sec. III, we describe the physical effects that control convexity in different models of self-limiting assembly. Here we see that the concentration dependence of the mean (or peak) self-limiting size, as well as the size dispersity, is controlled by a single combination of $\epsilon''_{\rm T}$ and $n_{\rm T}$.

The effect of finite convexity is to allow fluctuations in n around the peak size n_*. When $\epsilon''_{\rm T}$ and $n_{\rm T}$ are sufficiently large, the aggregate distribution follows a Gaussian

$$\rho_n(n \gg 1) \simeq e^{n_* \beta(\mu - \epsilon_*)} e^{-(n-n_*)^2/2\langle \Delta n^2 \rangle}, \tag{17}$$

where $\langle \Delta n^2 \rangle$ characterizes the variance of aggregate sizes relative to n_*. Assuming that the Gaussian distribution of aggregates is well separated from the monomer peak, the size fluctuations around n_* may be summed in $\phi_n = n\rho_n$, yielding the same mass-action formula as Eq. (11), but with a redefined CAC

$$\phi_* \approx \frac{e^{\beta\epsilon_*}}{(n_* \sqrt{2\pi\langle \Delta n^2 \rangle})^{1/n_*}} \quad \text{(Gaussian)}. \tag{18}$$

Compared to the two-state approximation, ϕ_* is depressed by a factor proportional to $\langle \Delta n^2 \rangle^{1/2n_*}$ due to the comparative increase in the number of aggregate states and associated entropy. Likewise, well above the CAC, the monomer population is depressed (relative to the two-state approximation) by the same factor. Combining this effect into the chemical potential with the peak aggregate condition in Eq. (12) gives the following prediction for the peak (mean) aggregate:

$$n_* \simeq n_{\rm T}\left(1+\frac{k_BT}{n_{\rm T}^3\epsilon''_{\rm T}}\ln\left[\frac{\Phi}{n_{\rm T}\sqrt{2\pi\langle\Delta n^2\rangle}}\right]\right) \quad \text{(Gaussian)}, \qquad (19)$$

where we are considering only the leading correction to $n_* - n_{\rm T}$. In this same limit, aggregate dispersity becomes

$$\frac{\langle\Delta n^2\rangle^{1/2}}{n_{\rm T}} \simeq \frac{1}{(n_{\rm T}^3\beta\epsilon''_{\rm T})^{1/2}} \quad \text{(Gaussian)}, \qquad (20)$$

where again averages are taken with respect to aggregate distribution ρ_n (i.e., the number average).

The results of the Gaussian approximation in Eqs. (19) and (20) highlight two physical effects of convexity. First, in Eq. (19), the mean aggregate size always falls *slightly below* the minimal-energy target size (i.e., because $\Phi < 1$, and hence the logarithmic factor is always negative). This "suboptimal" aggregate size derives from the translational entropic preference for smaller-n aggregates, and hence this weak depression of n_* decreases with increasing supersaturation as the translation free energy of aggregates $(k_BT/n_{\rm T})\ln(\Phi/n_{\rm T})$ tends toward zero. Second, the relative shift of mean aggregation number $(n_* - n_{\rm T})/n_{\rm T}$ and the relative size variance $\langle\Delta n^2\rangle^{1/2}/n_{\rm T}$ decrease with the reciprocal of $\epsilon''_{\rm T}n_{\rm T}^3/k_BT$. Hence, corrections from size variations become small, in relative terms, either for sharp minima when $\epsilon''_{\rm T} \gg k_BT$ or for larger aggregate number when $n_{\rm T} \gg (k_BT/\epsilon''_{\rm T})^{1/3}$. While at first glance this might suggest a generic tendency toward monodisperse aggregation in the large-$n_{\rm T}$ limit, SLA models described in Sec. III show convexity to be a *decreasing* function of $n_{\rm T}$. Hence, as it turns out the decrease of relative size fluctuations with target size becomes nontrivially dependent on the geometric sensitivity of aggregation energy.

c. Self-limitation without minima

Before considering landscapes with more complex equilibria, we note that it is possible to construct functional forms of $\epsilon(n)$ that exhibit self-limitation without local minima. As an example, consider a variation of the general type of energetics described in Sec. II.A.2,

$$\epsilon(n) = -\epsilon_0 + \Delta(n) \qquad (21)$$

where $\Delta(n)$ is a monotonically decreasing, and convex, function of n so that the minimal energy per particle occurs for infinite aggregates, i.e., $\epsilon(n\to\infty) = -\epsilon_0$. The condition for a maximum in ρ_n in Eq. (12) gives

$$\mu + \epsilon_0 \equiv -(\Delta\mu)_\infty = \Delta(n_*) + n_*\Delta'(n_*). \qquad (22)$$

Because the chemical potential is bounded from above by $-\epsilon_0$ [from Eq. (14)] and hence $(\Delta\mu)_\infty > 0$, the conditions for finite optimal aggregate size can be satisfied (at some accessible μ) provided that $\Delta(n)$ decreases *faster* than $1/n$ or, more specifically, that

$$\frac{\Delta'(n)}{\Delta(n)} < -\frac{1}{n} \qquad (23)$$

for some range of finite n. For example, for any model that approaches the bulk energy as a power law $\Delta(n) = \Delta_0/n^\gamma$, where $\gamma > 1$, it is straightforward to show that the peak size obeys $n_* = [(\gamma-1)\Delta_0/(\Delta\mu)_\infty]^{1/\beta}$, which increases continuously with concentration as $\Delta\mu_\infty \to 0$ from above.

While the previous argument shows that it is possible to construct mathematical examples of minima-free energy densities that result in finite-n peaks in ρ_n, physical cases of SLA fall outside of this category. For example, generic physical grounds suggest that aggregates possess a boundary, or surface, at which the assembly energy is different (generally higher) than in the interior, as described for cases of short-range cohesive interactions in Sec. II.A.2. In the limit of large n, the asymptotic contribution to the energy density from this boundary $n^{-1/d}$, where $d \geq 1$, will dominate over other possible terms falling off faster than n^{-1} [such as the $\Delta(n)$ term in Eqs. (21) and (23)], leading to unlimited assembly. Hence, we exclude such anomalous cases and focus the discussion on situations where self-limitation is directly associated with well-defined minima of $\epsilon(n)$.

3. Competing states of aggregation

Sections II.B.1 and II.B.2 describe the simplest case of self-limiting assembly: a concentration-controlled crossover, or "pseudotransition," from a monomer-dominated state to a state dominated by aggregates of one finite size. The finite aggregate size corresponds to a single minimum in $\epsilon(n)$, and the transition occurs at a single CAC. In this section, we overview the thermodynamics of cases in which assembly is characterized by multiple local minima, or instead by transitions between self-limiting and unlimited aggregation states. In these cases, the aggregation thermodynamics can exhibit a more complex dependence on concentration corresponding to *secondary CACs* between different aggregation states. While concentration-dependent transitions between different aggregates are commonly attributed to interactions between aggregates (Israelachvili, 2011), it is less widely appreciated that they can also occur in ideal aggregation models, which strictly neglect interaggregate interactions. As we review in Sec. III.A.2, the possibility of an "ideal" secondary CAC was first considered in the context of transitions between spherical and cylindrical surfactant micelles (Porte *et al.*, 1984; May and Ben-Shaul, 2001). In this section, we describe this behavior as a generic consequence of the translational entropy preferences for smaller aggregate sizes. As such, ideal secondary CACs can occur in a much broader class of SLA models.

a. Two finite aggregate states

We first describe a simple model with only two states of finite aggregates in equilibrium with a population of free monomers, for simplicity ignoring number fluctuations around these three states. We consider two states of *small* and *large* aggregates, corresponding to two well-defined local minima of $\epsilon(n)$, at $n_{\rm S}$ and $n_{\rm L} > n_{\rm S}$ subunits, respectively, as shown schematically in Fig. 6(a). In this case, aggregation is controlled not only by the difference in the respective energy minima $\epsilon_{\rm S}$ and $\epsilon_{\rm L}$ but also by the difference in the aggregation number. To understand aggregation in the presence of multiple minima, it is convenient to define the nominal CACs corresponding to either aggregate state, from Eq. (10),

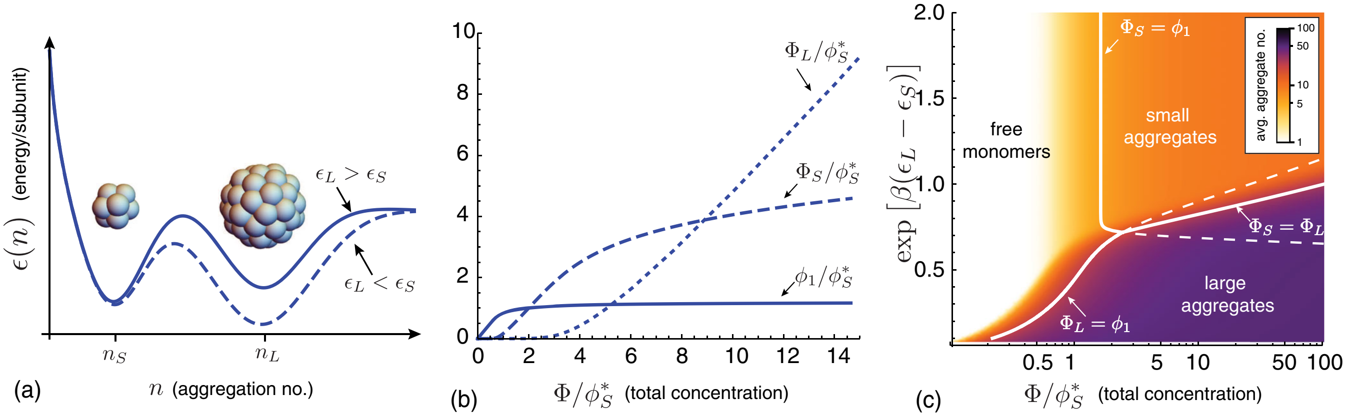

FIG. 6. (a) Schematic plot of the aggregation energetics for a case with two local minima corresponding to small and large finite aggregates containing n_S and n_L respective subunits in their target sizes. The solid curve shows a case where small aggregates are the global minimum of $\epsilon(n)$, while the dashed curve shows a case where the larger aggregate is the global minimum. The latter case can lead to secondary CAC transitions between small and large aggregates, as shown in (b), which plots the monomer, small aggregate, and large aggregate populations as functions of total concentration for large aggregates that are slightly energetically more favorable than small ones ($e^{\beta(\epsilon_\mathrm{L}-\epsilon_\mathrm{S})} = 0.8$). In this case, small aggregates dominate at intermediate concentrations but ultimately are overtaken by a second population of large aggregates above a second CAC. (c) Assembly state map for the two finite aggregate model, as a function of the per subunit energy difference between small and large aggregates and total concentration. The color scale shows the mean aggregate size, while the solid lines indicate boundaries between states dominated by monomers, n_S-mers, and n_L-mers. The dashed lines indicate where subdominant aggregates reach the monomer concentration. (b),(c) Results of the two finite aggregate model for $n_\mathrm{S} = 10$ and $n_\mathrm{L} = 50$.

$$\phi_\nu^* \equiv [n_\nu e^{-\beta n_\nu \epsilon_\nu}]^{-1/(n_\nu - 1)} \approx \frac{e^{\beta\epsilon_\nu}}{n_\nu^{1/n_\nu}} \quad \text{for } \nu = \mathrm{S}, \mathrm{L}. \tag{24}$$

These are concentrations at which aggregates of either type would overtake free monomers, were it not for the additional equilibrium between the populations of small and large aggregates. In terms of these concentration scales, the law of mass action takes the form

$$\Phi = \phi_1 + \phi_\mathrm{S}^* \left(\frac{\phi_1}{\phi_\mathrm{S}^*}\right)^{n_\mathrm{S}} + \phi_\mathrm{L}^* \left(\frac{\phi_1}{\phi_\mathrm{L}^*}\right)^{n_\mathrm{L}}, \tag{25}$$

where the last two terms represent the respective populations of subunits in n_S-mers and n_L-mers, which we denote as Φ_S and Φ_L. As shown previously for a single minimum in $\epsilon(n)$, in the limit of high concentration aggregation always proceeds toward the state with the lowest energy. However, in this case there are two possible thermodynamic scenarios for the concentration dependence, depending on the relative energy difference between small and large aggregates:

(i) $\epsilon_\mathrm{S} < \epsilon_\mathrm{L}$: Because $n_\mathrm{S} < n_\mathrm{L}$, in this case it is always true that $\phi_\mathrm{S}^* < \phi_\mathrm{L}^*$, which means that as concentration increases $\phi_1 \to \phi_\mathrm{S}^*$ before reaching ϕ_L^*. Above the threshold where $\phi_1 \approx \Phi_\mathrm{S}$, monomers remain effectively buffered at $\phi_1 \approx \phi_\mathrm{S}^*$, and it is straightforward to show that $\Phi_\mathrm{L} \ll \Phi_\mathrm{S}$.[6] Thus, when the smaller aggregate has a lower per subunit aggregation energy, aggregation proceeds as if there is only a single target state with a CAC at ϕ_S^* and never yields a significant number of large aggregates.

(ii) $\epsilon_\mathrm{S} > \epsilon_\mathrm{L}$: When the large aggregates are energetically favored, there are two possibilities. First consider the case of large energy differences, such that $e^{\beta(\epsilon_\mathrm{L}-\epsilon_\mathrm{S})} < n_\mathrm{L}^{1/n_\mathrm{L}}/n_\mathrm{S}^{1/n_\mathrm{S}}$. For this first regime $\phi_\mathrm{L}^* < \phi_\mathrm{S}^*$, and there is only a single CAC at the lower critical concentration for large aggregates. For the second regime, when the energy difference between large and small aggregates is smaller, in the range $1 > e^{\beta(\epsilon_\mathrm{L}-\epsilon_\mathrm{S})} > n_\mathrm{L}^{1/n_\mathrm{L}}/n_\mathrm{S}^{1/n_\mathrm{S}}$ the order of the CACs reverses, $\phi_\mathrm{S}^* < \phi_\mathrm{L}^*$, and leads to two CACs. As shown in Fig. 6(b), for this case, with increasing concentration from the dilute limit, the concentration reaches a first CAC at $\Phi \approx \phi_\mathrm{S}^*$, with a transition to a state dominated by small aggregates (i.e., $\Phi_\mathrm{S} \gg \phi_1, \Phi_\mathrm{L}$). This state persists until reaching a second CAC at $\Phi \approx \phi^{**}$, defined by a crossover in dominant aggregation state to $\Phi_\mathrm{L} > \Phi_\mathrm{S}$. The concentration threshold condition can be estimated by solving for the monomer concentration at which $\Phi_\mathrm{S}(\phi_1) = \Phi_\mathrm{L}(\phi_1) \gg \phi_1$ from Eq. (25):

$$\phi^{**} \approx \phi_\mathrm{L}^* \left(\frac{\phi_\mathrm{L}^*}{\phi_\mathrm{S}^*}\right)^{n_\mathrm{S}/(n_\mathrm{L}-n_\mathrm{S})}, \tag{26}$$

which is larger than the "bare" value ϕ_L^* owing to the depletion of free monomers by small aggregates and corresponds to a total concentration $\Phi^{**} \approx 2\phi^{**}$. The high-concentration regime above the second CAC is dominated by minimal-energy large aggregates but maintains a sizable amount of small aggregates (i.e., $\Phi_\mathrm{L} \gg \Phi_\mathrm{S} \gg \phi_1$) approximately buffered at the second CAC concentration.

---

[6] The assumptions of ideal aggregation require that Φ must remain below unity, a condition that requires $\phi_1 \ll \phi_\mathrm{L}^*$ for case (i).

To summarize this two-aggregate model, it is possible to have *two* pseudocritical transitions [as in Fig. 6(b)], first from disassembled monomers to small aggregates and then from small to large aggregates, provided that the energy difference between aggregates is sufficiently small, in the window

$$0 < \epsilon_{\mathrm{S}} - \epsilon_{\mathrm{L}} < \frac{k_B T}{n_{\mathrm{S}}} \ln n_{\mathrm{S}} - \frac{k_B T}{n_{\mathrm{L}}} \ln n_{\mathrm{L}}. \tag{27}$$

This is consistent with the secondary CAC behavior shown in the assembly state diagram in Fig. 6(c) calculated for the case of $n_{\mathrm{S}} = 10$ and $n_{\mathrm{L}} = 50$.

The physical origin of this double CAC behavior can be traced to a competition between the higher cohesive energy of large aggregates pitted against the higher per subunit translational entropy of smaller aggregates. This can be cast in terms of the chemical equilibrium between large and small aggregates, which requires

$$\mu = \epsilon_{\mathrm{S}} + \frac{k_B T}{n_{\mathrm{S}}} \ln \rho_{\mathrm{S}} = \epsilon_{\mathrm{L}} + \frac{k_B T}{n_{\mathrm{L}}} \ln \rho_{\mathrm{L}}. \tag{28}$$

Energetically favorable large aggregates ($\epsilon_{\mathrm{L}} < \epsilon_{\mathrm{S}}$) require aggregate concentrations to adjust to maintain a suitably higher translational entropy of small aggregates, namely, $(1/n_{\mathrm{S}}) \ln \rho_{\mathrm{S}} < (1/n_{\mathrm{L}}) \ln \rho_{\mathrm{L}}$, specifically

$$\rho_{\mathrm{S}}/\rho_{\mathrm{L}} = \frac{e^{n_{\mathrm{S}}\beta(\epsilon_{\mathrm{L}}-\epsilon_{\mathrm{S}})}}{\rho_{\mathrm{L}}^{(n_{\mathrm{L}}-n_{\mathrm{S}})/n_{\mathrm{L}}}}. \tag{29}$$

Equation (29) shows that the larger entropy of smaller aggregates requires that $\rho_{\mathrm{S}}/\rho_{\mathrm{L}} > 1$ provided that the concentration of large aggregates remains sufficiently small (below $\rho_{\mathrm{L}} < \rho_{\mathrm{L}}^{**} = [e^{\beta(\epsilon_{\mathrm{L}}-\epsilon_{\mathrm{S}})}]^{n_{\mathrm{S}} n_{\mathrm{L}}/(n_{\mathrm{L}}-n_{\mathrm{S}})}$). Hence, when n_{S} and the differential in aggregation energy are small enough, small aggregates remain more populous than large aggregates up to total concentrations that exceed the first CAC to the small aggregate state until the second CAC.

This simplified model illustrates a generic conclusion. Even if an aggregate state does not correspond to the *global* minimum of $\epsilon(n)$, it may exhibit an entropically stabilized window of thermodynamic dominance at intermediate concentrations provided that its target size is sufficiently small and its energy is sufficiently close to the global minimum. Next we illustrate this entropic stabilization of finite compact aggregates in models for which the competing states are *unlimited*.

b. Finite and unlimited aggregates

While polymorphic assembly into multiple finite-number aggregates occurs in some natural and biomimetic systems (Wingfield *et al.*, 1995; Sun *et al.*, 2007; Lutomski *et al.*, 2018) and may be desirable for nanomaterials applications, cases in which aggregates *change dimensionality* are more common: that is, aggregate structures that remain finite in at least one spatial direction but undergo essentially unlimited growth in other directions. The most common examples are amphiphilic assemblies, which can form spherical micelles (finite in all directions), cylindrical micelles (finite in two spatial dimensions, unlimited in one), or lamellar or layered assemblies (finite in one dimension, unlimited in two). In Sec. III.A.2 we describe the molecular ingredients that lead to polymorphic transitions between aggregate dimensionality based on a model of surfactant assembly (May and Ben-Shaul, 2001; Bergström, 2016). In this section, we illustrate how the principles of secondary CAC behavior apply to models that can exhibit states of finite aggregation number (such as spherical aggregates) that can transition to states of 1D aggregation or a bulk unlimited morphology.

Since our primary interest is to describe conditions where ideal aggregation gives rise to concentration-dependent transitions in morphology, we consider a minimal description of a generic model including finite and unlimited aggregation states. As summarized in Fig. 7(a), this model considers three disconnected "branches" of assembly:

1. *Finite compact aggregates* with respective target size and energy n_{F} and ϵ_{F}.
2. *1D aggregates* with energy per subunit

   $$\epsilon(n) = \epsilon_{\mathrm{1D}} + \Delta_0/n,$$

   where $\Delta_0 > 0$ characterizes the cost of finite end caps and $\epsilon_{\mathrm{1D}} < 0$ is the limiting $n \to \infty$ per subunit assembly energy in this morphology.
3. *Bulk aggregates* with energy density ϵ_{bulk}. Here we consider only one macroscopic aggregate ($n \to \infty$) with negligible boundary energy.

Based on the foregoing analysis of the two-finite assembly state model, it can be anticipated that secondary CAC behavior from finite to 1D aggregation takes places when the $n \to \infty$ energy density of 1D aggregates is lower than for finite aggregates, but the energy gap is sufficiently small that the translational entropy associated with the compact aggregates can stabilize a window of n_{F}-mer aggregates. Likewise, when the energy density of the bulk state falls below these dimensionally limited states, we anticipate an upper limit to concentration (i.e., saturation) that can maintain equilibrium with dispersed aggregates.

With this in mind, we consider a simplified law of mass action for subunit populations

$$\Phi = \phi_1 + \Phi_{\mathrm{F}} + \Phi_{\mathrm{1D}} + \Phi_{\mathrm{bulk}}, \tag{30}$$

where the terms, respectively, describe the populations of free monomers, subunits in a single finite aggregate size, 1D aggregates of various size, and bulk aggregation. The population of subunits in finite aggregates is given by (neglecting number fluctuations)

$$\Phi_{\mathrm{F}} = n_{\mathrm{F}} (\phi_1 e^{-\beta\epsilon_{\mathrm{F}}})^{n_{\mathrm{F}}}, \tag{31}$$

while the 1D aggregate population is given by

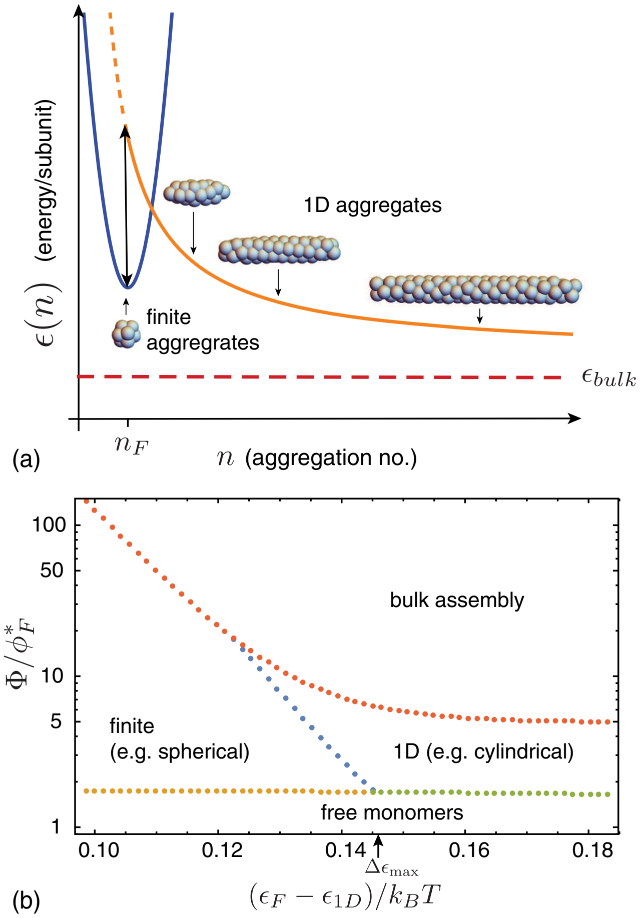

FIG. 7. (a) Schematic plot of polymorphic aggregation energetics with three competing branches of assembly: finite aggregates with a local minimum at n_{F}, 1D aggregates, and bulk aggregates ($n \to \infty$ energy shown as a horizontal dashed line). In this case, the infinite 1D aggregate has a lower per subunit energy than finite aggregates, and there is a barrier in total energy δ that separates these states at $n = n_{\mathrm{F}}$, i.e., the double arrow in (a) corresponds to δ/n_{F}. (b) Phase diagram for concentration-dependent size selection. The dominant aggregation state is shown for a system with coexistence among finite aggregates with $n_{\mathrm{F}} = 100$ subunits, separated by an energy gap $\epsilon_{\mathrm{F}} - \epsilon_{\mathrm{1D}}$ and a barrier of δ [Eq. (34)] to 1D aggregates. There is an additional per subunit energy gap of $\epsilon_{\mathrm{1D}} - \epsilon_{\mathrm{bulk}} = 0.0005 k_B T$ between the 1D and bulk aggregates. The horizontal axis gives the energy gap between spheres and cylinders, and the vertical axis gives the total concentration relative to the CAC for finite aggregates $\phi_{\mathrm{F}}^* \simeq e^{\beta\epsilon_{\mathrm{F}}}/n_{\mathrm{F}}$. The boundaries between monomers, finite aggregates, and 1D aggregates are determined by crossovers in the most populous aggregate type from Eq. (30), while the point of bulk saturation is determined by the point when $\mu = \epsilon_{\mathrm{bulk}}$. In this example, the energy per subunit in finite aggregates is fixed at $\epsilon_{\mathrm{F}} = -10\ k_B T$ and the end cap energy of the spherocylinders is $\Delta_0 = 20\ k_B T$. The maximum energy gap for which second CAC behavior occurs [$\Delta\epsilon_{\max} \approx 0.14 k_B T$; Eq. (35)] is indicated along the x axis.

$$\Phi_{\mathrm{1D}} \simeq \sum_{n=n_{\mathrm{F}}}^{\infty} n e^{-\beta\Delta_0} (\phi_1 e^{-\beta\epsilon_{\mathrm{1D}}})^n \simeq n_{\mathrm{F}} e^{-\beta\Delta_0} \frac{(\phi_1 e^{-\beta\epsilon_{\mathrm{1D}}})^{n_{\mathrm{F}}}}{1 - \phi_1 e^{-\beta\epsilon_{\mathrm{1D}}}}. \tag{32}$$

In Eq. (32), we made the additional assumptions that 1D assembly is not favorable below some aggregate size close to $n_{\mathrm{F}} > 1$, and that the monomer concentration remains well below the value $e^{\beta\epsilon_{\mathrm{1D}}}$ where Φ_{1D} diverges.[7] The results are not qualitatively sensitive to these approximations.

Based on these forms, it is straightforward to find the free monomer concentration ϕ_1^{**} where finite aggregates and 1D aggregates are equally populous, i.e., $\Phi_{\mathrm{F}}(\phi_1^{**}) = \Phi_{\mathrm{1D}}(\phi_1^{**})$,

$$\phi_1^{**} = e^{\beta\epsilon_{\mathrm{1D}}}(1 - e^{-\beta\delta}), \tag{33}$$

where

$$\delta = \Delta_0 + n_{\mathrm{F}}(\epsilon_{\mathrm{1D}} - \epsilon_{\mathrm{F}}) \tag{34}$$

is the energy difference, or "barrier," between an n_{F}-mer and a 1D aggregate of the same size; see Fig. 7(a). As previously described, a stable aggregate population requires at least a local minimum in the energy and hence a barrier necessarily separates aggregation states associated with distinct local maxima in population. The size of this barrier determines the window of secondary CAC transition behavior as follows.

First, note that $\delta > 0$ implies that the energy of forming two end caps on the 1D aggregate exceeds that of the target n_{F}-mer. Second, the existence of a second CAC requires this concentration to exceed the primary CAC to a n_{F}-mer dominated state, that is, the condition $\phi_1^{**} > \phi_{\mathrm{F}}^* \simeq e^{\beta\epsilon_{\mathrm{F}}}/n_{\mathrm{F}}^{1/n_{\mathrm{F}}}$. This gives an upper limit to the energy gap between n_{F}-mer aggregation and 1D assembly for second CAC behavior $\epsilon_{\mathrm{F}} - \epsilon_{\mathrm{1D}} < \Delta\epsilon_{\max}$, with

$$\Delta\epsilon_{\max} \approx \frac{k_B T}{n_{\mathrm{F}}} \ln n_{\mathrm{F}} + k_B T \ln(1 - e^{-\beta\delta}), \tag{35}$$

the stability window of the n_{F}-mer state expands, in terms of $\Delta\epsilon_{\max}$, with both decreasing target size and increasing energy barrier to 1D aggregation. For energy gaps larger than this limiting condition, the intermediate n_{F}-mer state disappears.[8]

Last, note that monomers reach chemical equilibrium with the bulk state at a concentration $\phi_1 = e^{\beta\epsilon_{\mathrm{bulk}}}$, which sets an additional condition for second CAC behavior $\phi_1^{**} < e^{\beta\epsilon_{\mathrm{bulk}}}$. This gives the upper limit to the energy gap between 1D and bulk assembly for second CAC behavior before saturation

$$\epsilon_{\mathrm{1D}} - \epsilon_{\mathrm{bulk}} < -k_B T \ln(1 - e^{-\beta\delta}). \tag{36}$$

Equation (36) shows that the concentration range of the 1D aggregate state diminishes with an increasing energy barrier between compact and 1D aggregates.

An example of an assembly state diagram for this model under conditions $\epsilon_{\mathrm{F}} > \epsilon_{\mathrm{1D}} > \epsilon_{\mathrm{bulk}}$ is shown in Fig. 7(b). The concentration-dependent state of aggregation is plotted versus

---

[7] Specifically, we assume $n_{\mathrm{F}} > e^{\beta\delta}$. For $n_{\mathrm{F}} < e^{\beta\delta}$, $\Phi_{\mathrm{1D}} \simeq e^{-\beta\Delta}(\phi_1 e^{-\beta\epsilon_{\mathrm{1D}}})^{n_{\mathrm{F}}+1}(1 - \phi_1 e^{-\beta\epsilon_{\mathrm{1D}}})^{-2}$, $\phi_1^{**} \approx e^{\beta\epsilon_{\mathrm{1D}}}(1 - e^{-\beta\delta/2}/n_{\mathrm{F}}^{1/2})$.

[8] Note that Eq. (35) is an implicit relation for $\Delta\epsilon_{\max}$ since δ is a function of $\epsilon_{\mathrm{1D}} - \epsilon_{\mathrm{F}}$. However, in the limit $\delta \gg 1$ the maximum gap is $\Delta\epsilon_{\max} \approx (\epsilon_{\mathrm{1D}} + k_B T \ln n_{\mathrm{F}})/n_{\mathrm{F}}$. Similarly, in the limit $n_{\mathrm{F}} < e^{\beta\delta}$, $\Delta\epsilon_{\max} \approx (\Delta_0 - 2k_B T)/n_{\mathrm{F}}$; see footnote 7.

the energy gap between $n_{\rm F}$-mers and 1D aggregates for variable $\epsilon_{\rm F}$ with fixed $n_{\rm F}$, Δ, $\epsilon_{\rm 1D}$, and $\epsilon_{\rm bulk}$. Under these conditions, when the aggregation energy of compact $n_{\rm F}$-mers is larger than, but sufficiently close to, that of infinite 1D aggregation, the system undergoes a sequence of concentration-driven transitions: first from monomers to finite aggregations, then to 1D aggregates, and finally to bulk unlimited assembly. In Sec. III, we revisit the possibility of multi-CAC behavior in the context of polymorphism of surfactant aggregates, focusing on its microscopic origin for this effect in terms of an underlying molecular model of aggregation.

III. MECHANISMS AND MODELS OF SELF-LIMITING ASSEMBLY

In Sec. II, we overviewed the basic thermodynamics of ideal aggregation and described some of the generic ingredients and outcomes of finite-size equilibria. We showed that the key ingredient is a per subunit aggregation free energy that has one or more local minima as a function of aggregation number (for finite-number aggregates), or as a function of the size of one or more spatial dimensions of the aggregate (for spatially finite aggregates like quasicylindrical or planar structures). In this section, we review four broad mechanisms of self-limiting assembly and a physical example of each mechanism. In each case, we first focus on the physical ingredients that give rise to the self-limiting aggregation energetics and then illustrate some of the implications of the generic phenomonology overviewed in Sec. II.

To organize the discussion, we divide these mechanisms into two broadly delineated classes: *self-closing assembly* and *open-boundary assembly*. These classes are distinguished by the presence or absence of an open boundary and gradients in intra-aggregate stress in the target assembly. Self-closing assembly describes aggregates in which the subunits, by and large, share the same cohesive environment of neighboring subunits and further adopt a common shape in the target assembly. In contrast, self-limitation of open-boundary assemblies requires gradients of the inter-subunit forces throughout the aggregate. Within the discussion, we highlight the distinct outcomes and potential trade-offs between these different mechanisms in terms of size selection.

We divide the remainder of Sec. III into two main parts, focusing, respectively, on self-closing and open-boundary assemblies. Following the introduction of each broader class of SLA, we further subdivide each into two subclasses representing four basic physical mechanisms of SLA. For each of the four basic mechanisms, we overview the applicable physical systems and then introduce a simple example model that captures the emergence of self-limited assembly.

A. Self-closing assembly

We define self-closing assembly (SCA) as a class of self-limiting assembly that achieves a finite target size, or finite target dimension, due to anisotropic binding between neighbors that leads to a preferred rotation of neighbor bonds. Such interactions generically arise when subunits are tapered or wedge shaped, such that cohesive bonding leads to a relative rotation of the axes of neighbor units; see Figs. 2(a) and 9(a). In combination with the relative displacement of subunit centers, this relative rotation, when built up over multiple subunits, leads to a preferred intra-assembly curvature along one or more principal directions. In the simplest case, this can be visualized as 1D loops of subunits, whose preferred curvature radius $R_{\rm close}$ leads the structure to close upon itself.

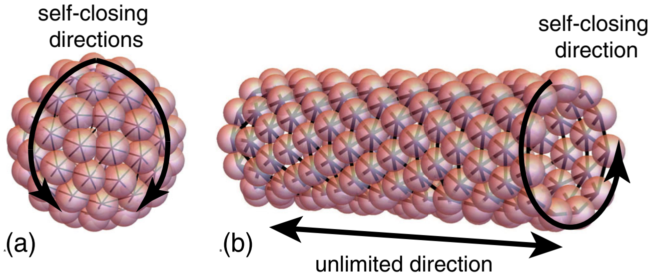

FIG. 8. (a) Schematic of spherical shell assembly with two independent self-closing directions of assembly. (b) Schematic of tubule assembly with one self-closing direction (circumferential) and one unlimited direction (axial) of assembly.

We include in the SCA class structures that close upon themselves in all directions of assembly and thus achieve a finite number of subunits, such as the spherical shell in Fig. 8(a), as well as structures that close in one or more directions but remain unlimited in others, such as the tubule in Fig. 8(b). In the latter case, structures have an unlimited number of subunits but achieve a finite size in the self-closing direction(s). For example, the tubule is unlimited in the axial direction but has a well-defined radius and corresponding number of subunits in the circumferential direction. The underlying principles for size selection remain the same as for finite number; namely, the self-limiting size W of a self-closing direction is determined by the minimum of the energy per subunit with respect to W. This can be readily understood by considering a system with nearly all of its N subunits assembled in aggregates of unlimited number, and correspondingly a negligible fraction of free monomers. To first approximation, the energetic costs of free edges and the translational entropy of the unlimited aggregates can be neglected, and the concentration of free monomers can be assumed to be a small contribution to the total free energy $F \approx N\epsilon(W)$. Hence, the thermodynamic equilibrium at fixed N corresponds to the selection of the energy-minimizing size W_*, which corresponds to $\partial_W \epsilon|_{W_*} = 0$. We describe similar considerations for self-limiting, open-boundary assemblies in Sec. III.B.1.

Strictly speaking, for a SCA it need not be necessary to identify a continuous loop of bonds along the self-closing direction(s) or that target curvatures are strictly uniform. We require only that there are one or more periodic directions on a representative 2D surface of the aggregate (such as the surface spanned by the subunit centers). Moreover, the preferred curvature does not need to select a perfectly commensurate number of subunits per cycle, since physical subunits generically possess some flexibility of shape and cohesion (bonding) that permits fluctuation in the interunit rotation. In the simplest case, such as with fluidlike intra-assembly order, a SCA may accommodate such strains through uniform deformation of subunits and their bonds. However, certain physical examples introduce extra geometric constraints (such as solidlike, spherical shells) that require at least some variable intra-assembly strains. Notwithstanding the possibility of such gradients, we categorize an assembly as self-closing if its target size is selected through the curvature radius R_{close}, as opposed to the accumulation of stress gradients, which is later described as a distinct class of self-limiting assembly.

Physical examples of SCA can be divided roughly into two groups according to the ratio of subunit size, characterized by some thickness d, and the target curvature radius R_{close}. When $R_{\text{close}}/d \gg 1$ the target number of units per cycle is proportionately large. This case describes tubules, shells, and capsules. The second case $R_{\text{close}} \approx d$ describes assemblies whose curvature (and thus self-limited size) is most often selected and regulated by the molecular dimension itself, which is characteristic of amphiphiles and their micellar aggregates.

1. Shells, capsules, and tubules

We first review the case of tubule or shell-like assemblies. Examples of these are common in biology, where tapered protein subunits select a preferred radius of assembly curvature (Oosawa and Asakura, 1975). Quasicylindrical (tubular) examples include microtubules (Nogales, 2000; Cheng, Aggarwal, and Stevens, 2012) and the bacterial flagella (Namba and Vonderviszt, 1997), while quasispherical (shell and capsule) examples include clathrin cages (Kirchhausen, Owen, and Harrison, 2014; Giani *et al.*, 2017; Bucher *et al.*, 2018; Mettlen *et al.*, 2018), viral capsids (Zlotnick and Mukhopadhyay, 2011; Mateu, 2013; Hagan, 2014; Bruinsma and Klug, 2015; Perlmutter and Hagan, 2015; Hagan and Zandi, 2016; Twarock *et al.*, 2018; Zandi *et al.*, 2020), bacterial microcompartments (Schmid *et al.*, 2006; Iancu *et al.*, 2007; Kerfeld, Heinhorst, and Cannon, 2010; Rae *et al.*, 2013; Chowdhury *et al.*, 2014; Bobik, Lehman, and Yeates, 2015; Kerfeld and Melnicki, 2016), and other protein-shell organelles (Sutter *et al.*, 2008; Pfeifer, 2012; Nott *et al.*, 2015; Zaslavsky *et al.*, 2018). Recently, engineered tapered or patchy colloids have also drawn interest for their ability to realize synthetic analogs of self-closing tubules and shells (Li, Josephson, and Stein, 2011; Morphew and Chakrabarti, 2017), although many if not most such realizations to date are properly categorized as analogs to micelles where curvatures are comparable to colloidal dimension. Whatever the underlying subunit structure, the existence of 1D curvature ensures equilibrium self-limitation along only one of the two assembly dimensions in the tubular constructs, while the preferred positive Gaussian curvature of shells and capsules leads to equilibrium states with finite subunit number.

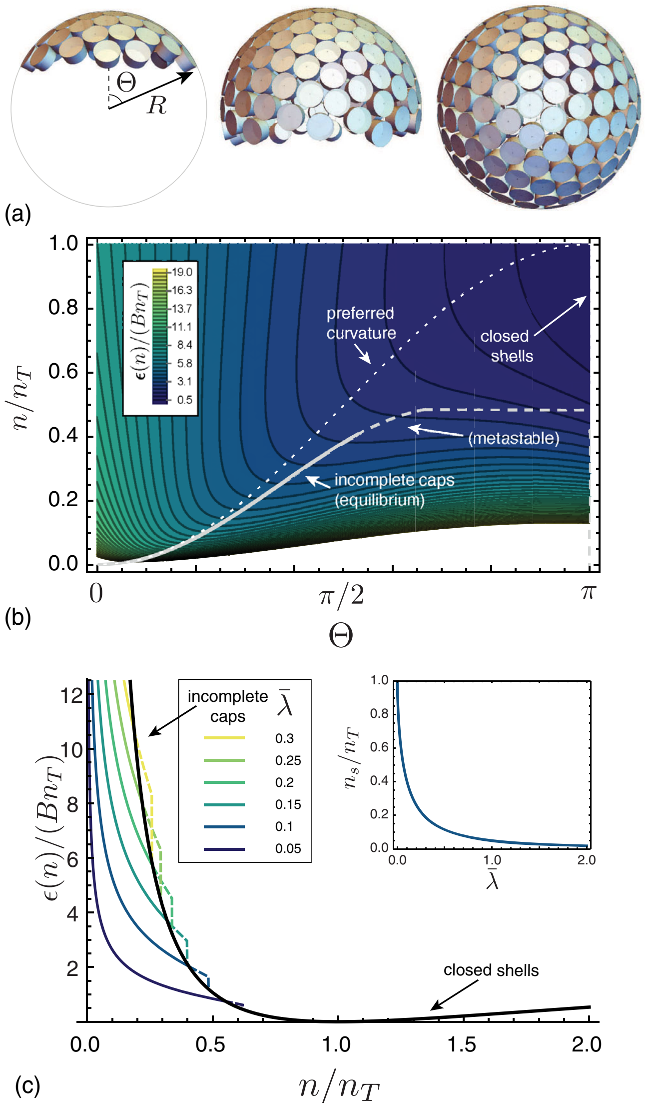

FIG. 9. (a) Schematic geometry of a "fluid capsule" of tapered subunits that assumes partial shell geometries of spherical caps with aperture angle $0 < \Theta \leq \pi$. (b) Contour map of the energy density landscape of the capsule model as a function of aperture angle and the ratio of subunits n to the preferred number in the ideal closed shell n_{T}. Low (high) values of $\epsilon(n)$ appear in purple (red). The dotted line shows a partial shell with the target curvature, while the solid line indicates the minimal-energy cap, whose curvature radius is slightly compressed by the line tension of the boundary. Beyond a threshold cap size $n = n_{\text{S}} \simeq 0.4 n_{\text{T}}$, this open cap becomes unstable to preclosure and the minimal-energy branch runs along $\Theta \to \pi$. This landscape corresponds to a dimensionless line tension $\bar{\lambda} = 0.1$. (c) Plots of the minimal energy branches of assembly. Incomplete caps are shown as colored curves (metastable portions are dashed), and the closed shell ($\Theta \to \pi$) is shown as a black curve. Inset: the size n_{S} corresponding to the preclosure, or "snap," transition between stable open caps and closed shells as a function of line tension.

To illustrate the self-limiting thermodynamics of shells and capsules, where target curvature radii are much larger subunit dimensions, consider the following minimal model. Spherical fluid capsules, shown schematically in Fig. 9(a), are composed of subunits with nominal area a_0 and a tapered shape that favors a preferred target spherical curvature radius R_{T}. Here we restrict the analysis to cases in which the curvature preference sufficiently disfavors locally anisotropic curvature (Lázaro, Dragnea, and Hagan, 2018) to limit incomplete assembly to caplike aggregation states with positive Gaussian curvature. Competition between incomplete assemblies with positive and zero Gaussian curvature was recently considered by Mendoza and Reguera (2020). We assume that the cap covers an axisymmetric "cap" domain of radius R from its pole up to the aperture angle Θ; see Fig. 9(a). While a closed capsule with a preferred curvature R_{T} has a target aggregation number $n_{\mathrm{T}} = 4\pi R_{\mathrm{T}}^2/a_0$, a capsule may realize a different aggregation number $n = 2\pi R^2(1 - \cos\Theta)$ provided that it is either open (i.e., $\Theta \neq \pi$) or deformed from its preferred taper (i.e., $R \neq R_{\mathrm{T}}$).

Taking the simplest possible model, we assume that the intrashell order is fluidlike, such that the only elastic penalty derives from bending deformations away from target curvature, which we consider via a membranelike bending energy

$$E_{\mathrm{bend}} = \frac{B}{2}\int dA \left(\frac{1}{R} - \frac{1}{R_{\mathrm{T}}}\right)^2, \tag{37}$$

where B is a bending modulus and the area integration is carried out over the incomplete shell. Additionally, we consider the line energy associated with the open edge for an incomplete cap

$$E_{\mathrm{open}} = 2\pi R \sin\Theta \lambda, \tag{38}$$

where λ is the energy per unit length of the exposed edge, associated with fewer cohesive bonds as well as a difference in solvation of subunits at the edge. The aggregation energy as a function of n then takes the form

$$\begin{aligned}\epsilon(n,\Theta) = {}& -\epsilon_{\mathrm{T}} + \frac{2\pi B}{n_{\mathrm{T}}}\left(\sqrt{\frac{n_{\mathrm{T}}(1-\cos\Theta)}{2n}} - 1\right)^2 \\ & + \frac{\sqrt{2\pi a_0}\lambda}{n^{1/2}}\frac{\sin\Theta}{\sqrt{1-\cos\Theta}},\end{aligned} \tag{39}$$

where $-\epsilon_{\mathrm{T}}$ is the bare aggregation energy for subunits in the bulk of undeformed capsules.

The form of $\epsilon(n,\Theta)$ in Eq. (39) has a global minimum for closed capsules of the target size (i.e., $n = n_{\mathrm{T}}$ and $\Theta = \pi$). Nevertheless, the combination of bending elasticity and the edge energy of incomplete shells influences assembly for $n \neq n_{\mathrm{T}}$. This can be seen by plotting the landscape of assembly energetics in the n/n_{T}–Θ plane, as shown in Fig. 9(b). For small aggregate sizes $n/n_{\mathrm{T}} \ll 1$, caps lock into the preferred curvature $R \to R_{\mathrm{T}}$ described by the condition $\cos\Theta \to 1 - 2n/n_{\mathrm{T}}$, which is shown as a dotted line in Fig. 9(b). As n increases, the edge energy favors compression of the open caps to smaller curvature radii $R < R_{\mathrm{T}}$. The amount of this shape compression grows up to a critical aggregation number n_{S}, beyond which the minimal-energy capsule snaps discontinuously to a closed shell of suboptimal size, that is, $\Theta \to \pi$ for $n_{\mathrm{S}} < n < n_0$.

Hence, the minimal aggregration versus n generically exhibits two branches [Fig. 9(c)]: an open cap for $n < n_{\mathrm{S}}$ and an edge-free closed shell for $n_{\mathrm{S}} < n < n_{\mathrm{T}}$. The transition between these two branches can be understood in terms of "nucleation" of an open pore in an overcurved shell, where $n_{\mathrm{T}} - n_{\mathrm{S}}$ corresponds to the size of the critical nucleus. The inset of Fig. 9C shows that the critical preclosure size n_{S} generically decreases with an increasing dimensionless ratio of edge energy and bending stiffness $\bar{\lambda} \equiv (a_0 n_{\mathrm{T}}/8\pi)^{1/2}\lambda/B$.

Notwithstanding the generic existence of a transition between open-cap and preclosed branches of the energy landscape, the transition does not lead to stable partial shells [i.e., a minima in $\epsilon(n)$ for $n \neq n_{\mathrm{T}}$]. Hence, while the open-cap branch and its transition to the closed shell may have implications for assembly pathways and kinetics, the equilibrium distribution of self-limiting capsules is independent of the edge energy and generically governed by the energetics of the closed-shell branch. This fact has further generic consequences for the concentration dependence and dispersity of aggregate size, both of which are governed by the product of the convexity and cube of the target size $\epsilon''(n_{\mathrm{T}})n_{\mathrm{T}}^3/k_B T$ according to Eqs. (19) and (20). The $\Theta \to \pi$ limit of Eq. (39) shows that $\epsilon''(n_{\mathrm{T}}) = B\pi/n_{\mathrm{T}}^3$. Hence, the decrease of convexity with target assembly size precisely cancels that entropic factor of n_{T}^3, such that the relative shift in mean aggregate size and relative dispersity, $(n_* - n_{\mathrm{T}})/n_{\mathrm{T}}$ and $\langle \Delta n^2\rangle^{1/2}/n_{\mathrm{T}}$, respectively, are limited only by the ratio of bending modulus to thermal energy $B/k_B T$ and are independent of self-closing target size.[9] Therefore, to regulate the self-limiting size of self-closing assemblies in absolute terms, the rigidity of subunits and their angular interactions must grow with target size.

The predicted growth of size fluctuations with target size would seem to contradict observations of the best studied example of self-closing shells, icosahedral virus capsids. At conditions of optimal assembly, size polydispersity of viruses is remarkably small. In fact, this high degree of monodispersity has been exploited by using 3D crystalline arrays of virus capsids for optical applications requiring precise spatial periodicity (Young *et al.*, 2008; Dang *et al.*, 2011; Minten *et al.*, 2011; Steinmetz *et al.*, 2011; Judd *et al.*, 2014; Malyutin *et al.*, 2015; Chen *et al.*, 2016; Delalande *et al.*, 2016; Park *et al.*, 2016; Rother *et al.*, 2016; Brillault *et al.*, 2017). While high-precision measurements of size dispersity in capsids are challenging, electron microscopy structures that were obtained without the assumption of icosahedral symmetry show that as many as 40% of alphavirus nucleocapsid core particles exhibit defects (Wang *et al.*, 2015; Wang, Mukhopadhyay, and Zlotnick, 2018), and hence some dispersity in shape. Some nonicosahedral structures, like immature human immunodeficiency virus (HIV) caspids, exhibit variations in the number of subunits (~1000 GAG protein subunits) that are of the order of the mean capsid size (Briggs *et al.*, 2009), although such effects may also be attributed to assembly kinetics (Dharmavaram *et al.*, 2019). More recently size distributions of hepatitis B virus (HBV) capsids (and capsids of other viruses) have been achieved at or near single-subunit precision using resistive pulse sensing (Zhou *et al.*, 2011, 2018), mass spectrometry (Uetrecht *et al.*, 2011), and charge detection mass spectrometry (Pierson *et al.*, 2014, 2016; Lutomski *et al.*, 2018). Although metastable defective capsids have been observed (Pierson *et al.*, 2016; Lutomski

---

[9] This continues until the asymptotic limit of zero spontaneous curvature ($n_{\mathrm{T}} \to \infty$), at which point the free energy per subunit is independent of size and the size distribution becomes an unlimited exponential (Helfrich, 1986), as shown for 1D assemblies in Sec. II.A.2.

et al., 2018), these measurements show that at long times (potentially corresponding to a near equilibrium state) the population is dominated by icosahedral capsids with the native size.[10]

To place these measurements in the context of the results for the fluid shell model, we note that bending moduli for virus capsids have been estimated from the force-displacement curves measured in nanoindentation experiments in which virus capsids are compressed using an AFM tip.[11] Estimated bending modulus values vary from $10k_BT$ to $200k_BT$, but typical values fall in the higher end of the range $[(100–200)k_BT]$. [For comparison, bending moduli of lipid bilayer membranes are typically in the range $(10–20)k_BT$.] Using the previous result that $\langle\Delta n^2\rangle^{1/2} \approx n_0/\sqrt{\pi B/k_BT}$, with $n_0 = 120$ and an estimate of $B \sim 60k_BT$ for HBV,[12] gives a root mean squared size fluctuation of about nine dimers, considerably larger than the long-time experimental estimates.

This discrepancy highlights two physical ingredients neglected in the previously described model: the discrete subunit size and the fact that most virus capsids, as well as most other protein shells, are crystalline rather than fluid. As discussed in Sec. V.B, the arrangement of proteins within icosahedral capsids can be mapped onto a triangular net. However, tiling a spherical topology with a triangular lattice requires the formation of 12 fivefold sites, often considered defects in a hexagonal packing. A number of equilibrium calculations have shown that the elastic energy of the defects themselves and interdefect elastic interactions significantly affect the energy landscape of such shells (Bruinsma et al., 2003; Zandi et al., 2004; Chen, Zhang, and Glotzer, 2007; Li et al., 2018, 2019; Mendoza and Reguera, 2020). These effects are reflected in local minima in the aggregate energy at certain "magic numbers" of subunits (Zandi et al., 2004). These minima correspond to shells with high degrees of symmetry, with the in-plane bonding energies corresponding to shells with icosahedral symmetry. Thus, when including the energetics of this bond ordering a size fluctuation of even one subunit can incur a significant energy cost ($\gtrsim 10k_BT$) since it requires disruption of the low-energy geometry, for example, through the introduction of a pair of fivefold and sevenfold defects. In fact, the metastable structures observed in HBV capsids are typically found at discrete intervals corresponding to deviations of multiple subunits from the native capsid size (Pierson et al., 2014, 2016; Lutomski et al., 2018), suggesting that typical fluctuations correspond to insertion or deletion of multisubunit oligomers (such as hexamers of the capsid protein), which would minimize disruptions to the capsid symmetry. For example, Figs. 10(a) and 10(b) show cryo-electron microscopy (cryo-EM) images and examples of 2D class averages, respectively, of *in vitro* assembly products of woodchuck hepatitis B virus (WHV) capsid proteins, which exhibit heterogeneous structures including icosahedral capsids, elongated closed shells, and shells with overlapping edges. Figure 10(c) shows a hypothetical model of a non-icosahedral closed shell in which insertion of additional hexamers leads to a prolate structure.

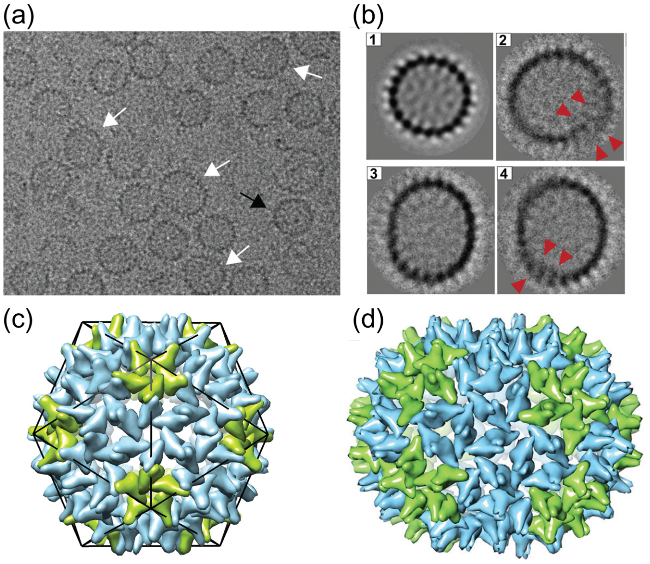

FIG. 10. Examples of well-formed and defective woodchuck hepatitis B virus (WHV) capsids. (a) Portion of a representative cryo-EM image of empty WHV capsids assembled *in vitro* in the absence of RNA (from capsid proteins with the C-terminal RNA binding deleted, wCp149). The population of capsids is structurally heterogeneous; the black arrow indicates an example of a $T = 4$ icosahedral capsid, while the white arrows indicate examples of defective capsids. (b) 2D class averages of (1) icosahedral and (2)–(4) defective particles. The red arrows in (2) and (4) indicate locations where the capsid shell is overgrown and overlaps itself. (c) Schematic model of a $T = 4$ icosahedral HBV capsid, with the monomers forming the 12 fivefold vertices colored green, and the others colored blue. (d) Hypothetical model of the structure of a capsid that is nonicosahedral but closed, with an elongated structure containing 150 protein dimers (the icosahedral capsid has 120 dimers). Here the green dimers are in pentamers or extend between pentamers and hexamers, and the blue dimers are in hexamers or extend between two hexamers. Adapted from Pierson et al., 2016.

Computational models of icosahedral assembly have also identified ensembles of defective capsules in which additional hexamers were added to the icosahedral shell (Nguyen, Reddy, and Brooks, 2009; Elrad and Hagan, 2010), although

---

[10] In fact HBV is dimorphic. Both *in vitro* and *in vivo* HBV capsid assembly yields mostly 120 protein dimer capsids with $T = 4$ icosahedral symmetry in the Caspar-Klug nomenclature (see Sec. V.B), but also a few percent of $T = 3$ icosahedral capsids. However, size fluctuations around the dominant $T = 4$ population were shown to be insignificant at long times.

[11] Note that estimating elastic moduli from the force-displacement curve is sensitive to the value chosen for thickness of the capsid shell, and the relationship between the atomic structure and the effective mechanical thickness remains at least somewhat obscure (May et al., 2011; May and Brooks, 2011).

[12] The bending modulus is calculated from the 3D Young's modulus $E = 0.26$ GPa obtained from nanoindentation measurements by Roos, Bruinsma, and Wuite (2010), and while assuming a thin shell model so that the 2D Young's modulus and bending modulus are, respectively, given by $Y = Et$ and $B = Yt^3/12(1 - \nu^2)$, with $t = 2.1$ nm the effective shell thickness (Wynne, Crowther, and Leslie, 1999; Roos, Bruinsma, and Wuite, 2010) and $\nu = 0.4$ the Poisson ratio (Uetrecht et al., 2008; Roos and Wuite, 2009).

these models and the corresponding defective structures differ from the HBV system. Hence, for more realistic models that incorporate both bending and bond-network elasticity we expect that these corrugations in the energy landscape (i.e., due to communicability with icosahedral symmetry) versus n will be superposed on the smooth landscapes illustrated in Fig. 9 for the fluid shell model, which may account for size fluctuations to be restricted for a limited set of low-energy values of n at or near high symmetry, or magic number capsomer arrangements.

2. Amphiphilic aggregates

Arguably the most common and well-studied class of self-limiting assemblies is amphiphiles. In the broadest sense, these refer to subunits with chemically dissimilar ends, which consequently favor distinct solvent environments. For example, lipids and surfactants possess oily hydrocarbon tails attached to a polar or charged head group (Israelachvili, 2011), which imparts a respective hydrophobic or hydrophilic character to either end of the same molecule [Fig. 11(a)]. Dispersing such amphiphiles in a solvent that has higher affinity to one end of the molecule generically drives them to form aggregates that partially hide, or sequester, the solvophobic portions while maintaining exposure of the solvophilic portions. Examples of such aggregrates, spherical and cylindrical micelles, and bilayer sheets are shown schematically in Figs. 11(b) and 11(c). These structures curve upon themselves, but do so on a length scale that is limited by, and comparable to, the size of the amphiphile itself, such as the molecular tail length in Fig. 11(a). The tendency to exclude unfavorable solvent from the core of the aggregate, in combination with the packing constraints of filling this region with the solvophobic portions, requires each amphiphilic subunit to span the entire thickness of the aggregate, which is fundamental to their self-limiting assembly.

In this section we describe a simple model to capture the self-limiting assembling of amphiphiles and highlight, in particular, how thermodynamic considerations of changes in aggregate thickness shape the preferred aggregate curvature but also give rise to polymorphism in the dimensionality of aggregates (spheres, cylinders, membranes, etc.). For illustration, we review a model for the thermodynamics of surfactant aggregation, of the type shown in Fig. 11(a), capturing central ingredients of the well-known packing model developed by Israelachvili, Mitchell, and Ninham (1976) and Israelachvili (2011). While this model aims to capture molecular elements of low-molecular weight surfactants and lipids, the essential thermodynamic features carry over to other amphiphilic assemblies, such as block copolymers in selective solvents (Leibler, Orland, and Wheeler, 1983; Halperin, Tirrell, and Lodge, 1992; Zhang and Eisenberg, 1996; Jain and Bates, 2003). The thermodynamics of amphiphile aggregation incorporates three ingredients: (i) the thermodynamics of area per solvophilic head group, (ii) the thermodynamics of molecular extension, and (iii) the constraint of uniform density in the solvophobic core, which links the first two elements.

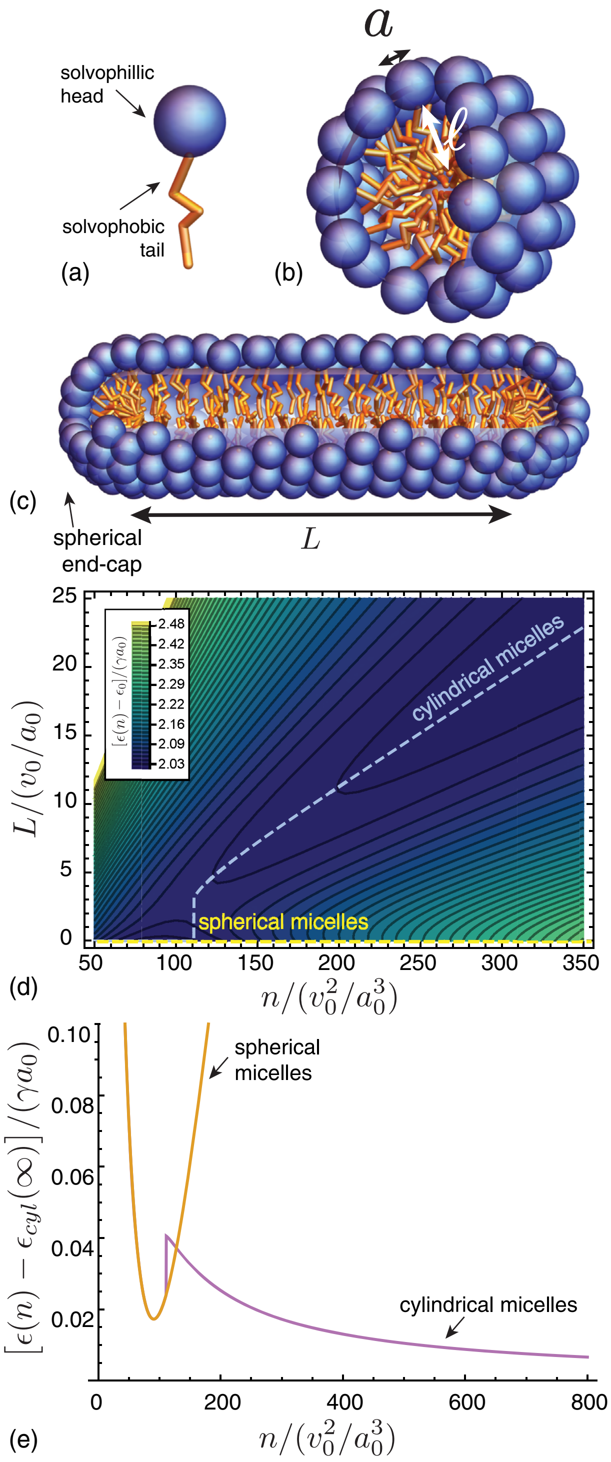

FIG. 11. (a) Sketch of a single-tailed surfactant molecule. (b) Schematic of a spherical micelle with the thickness of the solvophobic core ℓ and the area per head group a highlighted. A wedgelike portion of the micelle is cut away to illustrate the interior packing of tails. (c) Spherocylinder model of a wormlike micelle, which has a cylindrical portion of length L capped by two hemispherical micelles of equal radius. (d) Energy landscape for spherocylinders described by the model in Eq. (44) with $P^{-1}=2.25$ and $\bar{k}=0.1$. The dashed orange line shows the $L=0$ spherical micelle branch and the dashed pink line shows the $L=\infty$ wormlike micelle branch, whose corresponding energy as a function of reduced aggregation number is plotted in (e).

A simple model to describe the head-group energetics considers the per subunit energy to form aggregates with an area a at the solvent-core interface (Israelachvili, Mitchell, and Ninham, 1976; Israelachvili, 2011). The generic tendency to "hide" amphiphiles from the solvent is parametrized by a surface energy cost γa, where $\gamma > 0$ favors dense lateral packing of head groups. Competing against this lateral compression is the cost of interunit repulsive interactions, which in the simplest case are described by the two-body term in the virial series, giving a per subunit energy A_2/a, where $A_2 > 0$ (Tanford, 1974). These two terms can be combined into a single form

$$\epsilon_{\mathrm{int}} = \gamma\left(a + \frac{a_0^2}{a}\right), \tag{40}$$

where $a_0 = \sqrt{A_2/\gamma}$ is the optimal head-group area.

In combination with the tendency to achieve the optimal head-group area are additional thermodynamics of tail packing in the core, and the costs to extend its length ℓ. There are various models proposed for this effect, including a finite-maximum extension (Israelachvili, Mitchell, and Ninham, 1976; Israelachvili, 2011) or instead treating the core as a melt of flexible polymers (Dill and Flory, 1980; Ben-Shaul, Szleifer, and Gelbart, 1984; Nagarajan and Ruckenstein, 1991; Nagarajan, 2002). Here we adopt a simplified model used by May and Ben-Shaul for the free energy of tail length ℓ that spans from the solvent-core interface into the center of the aggregate

$$\epsilon_{\mathrm{stretch}} = \frac{k}{2}(\ell - \ell_0)^2, \tag{41}$$

where k is an elastic constant for intrasubunit stretch and ℓ_0 is a preferred length, which parametrizes the free energy cost of deformations from a preferred conformation state of the short tail. We consider k and ℓ_0 as a minimal description of the extensional thermodynamics and, like γ and a_0, these parameters can be varied through a combination of subunit structure and physical-chemical conditions, such as temperature and solvent properties.

Extensional energetics are linked by packing constraints associated with occupying the core with a fixed density of solvophobic portions of the subunits (Israelachvili, Mitchell, and Ninham, 1976; Israelachvili, 2011). These constraints vary with the dimensionality of the limited directions in the aggregate: $d_{\mathrm{L}} = 3$, spherical micelles; $d_{\mathrm{L}} = 2$, cylindrical micelles; and $d_{\mathrm{L}} = 1$, planar bilayers.[13] By considering the ratio between the core volume and interfacial area of an aggregate of thickness (radius) ℓ, it is straightforward to show that the solvophobic volume per subunit satisfies

$$v_0 = \frac{a\ell}{d_{\mathrm{L}}}. \tag{42}$$

As aggregates change their shape and number, uniform density requires adjustment of a and ℓ to maintain constant v_0. Using this constraint, we rewrite the assembly energy in terms of a single dimensionless thickness

$$r \equiv \frac{\ell}{(v_0/a_0)} = d_{\mathrm{L}}\frac{a_0}{a} \tag{43}$$

giving

$$\frac{\epsilon(r, d_{\mathrm{L}})}{\gamma a_0} = \left(\frac{d_{\mathrm{L}}}{r} + \frac{r}{d_{\mathrm{L}}}\right) + \frac{\bar{k}}{2}(r - P^{-1})^2 + \epsilon_0, \tag{44}$$

where $\bar{k} = k v_0^2/(\gamma a_0^3)$ is a scaled stretch modulus of the tail, and ϵ_0 parametrizes the negative energetic gain to assemble. The parameter P was introduced by Israelachvili as the packing parameter, a measure of the commensurability of the preferred shape with accessible aggregate geometries,

$$P \equiv \frac{a_0 \ell_0}{v_0}. \tag{45}$$

Written in this way, it is straightforward to understand how area and stretch thermodynamics compete to determine the optimal aggregate morphology. While area terms favor a thickness $r = d_{\mathrm{L}}$, stretch thermodynamics favors a thickness $r = 1/P$. Only for particular preferred head-group areas and subunit lengths do these two values coincide, i.e., when $P = 1/d_{\mathrm{L}}$; otherwise there is at least some shape frustration between these terms. As a heuristic, we therefore expect aggregation to favor the dimensionality d_{L} closest to $1/P$, which corresponds to the tapered geometry that most closely approximates the favored areal and thickness packing at uniform density.

A more complete picture of the polymorphism of aggregates is given by considering assembly landscapes that allow for transitions of micellar dimension. Figure 11(c) shows the structure of a "wormlike" micelle, modeled as a spherocylinder composed of a length L of cylindrical micelle capped by two equal-radius hemispherical micelle caps. As L increases, the fraction of the aggregate in the $d_{\mathrm{L}} = 2$ (versus $d_{\mathrm{L}} = 3$) packing increases, and thus consideration of the energy as a function of both n and L illustrates the landscape of aggregates intermediate to a uniformly cylindrical or spherical geometry. An example landscape is shown in Fig. 11(d), for a packing parameter $P^{-1} = 2.25$ intermediate to spheres and cylinders, which exhibits two branches of local minima. In the spherical branch, $L = 0$ and changes in number are accommodated purely through changes in micelle radius. The second, cylindrical branch appears only above a threshold aggregate number ($n \simeq 111 v_0^2/a_0^3$), beyond which further subunit addition is accommodated by increasing the length. The per subunit aggregation energies for these two branches are shown in Fig. 11(e). An energy barrier separates the convex minimum of the spherical branch from the cylindrical branch, which asymptotically approaches the global minimum of ϵ at $n \to \infty$ via the $1/n$ falloff characteristic of 1D assembly. The origin of this barrier between spherical and cylindrical aggregates derives from the fact that the preferred radii of these two micelle types are different, and hence the confinement energy of the end caps on the spherocylindrical micelle exceeds that of a minimal-energy

---

[13] Note that d_{L} refers to the number of limited directions, while we use d to refer to the dimensionality of the unlimited directions, e.g., $d = 1$ and $d = 2$ for cylindrical and lamellar aggregates.

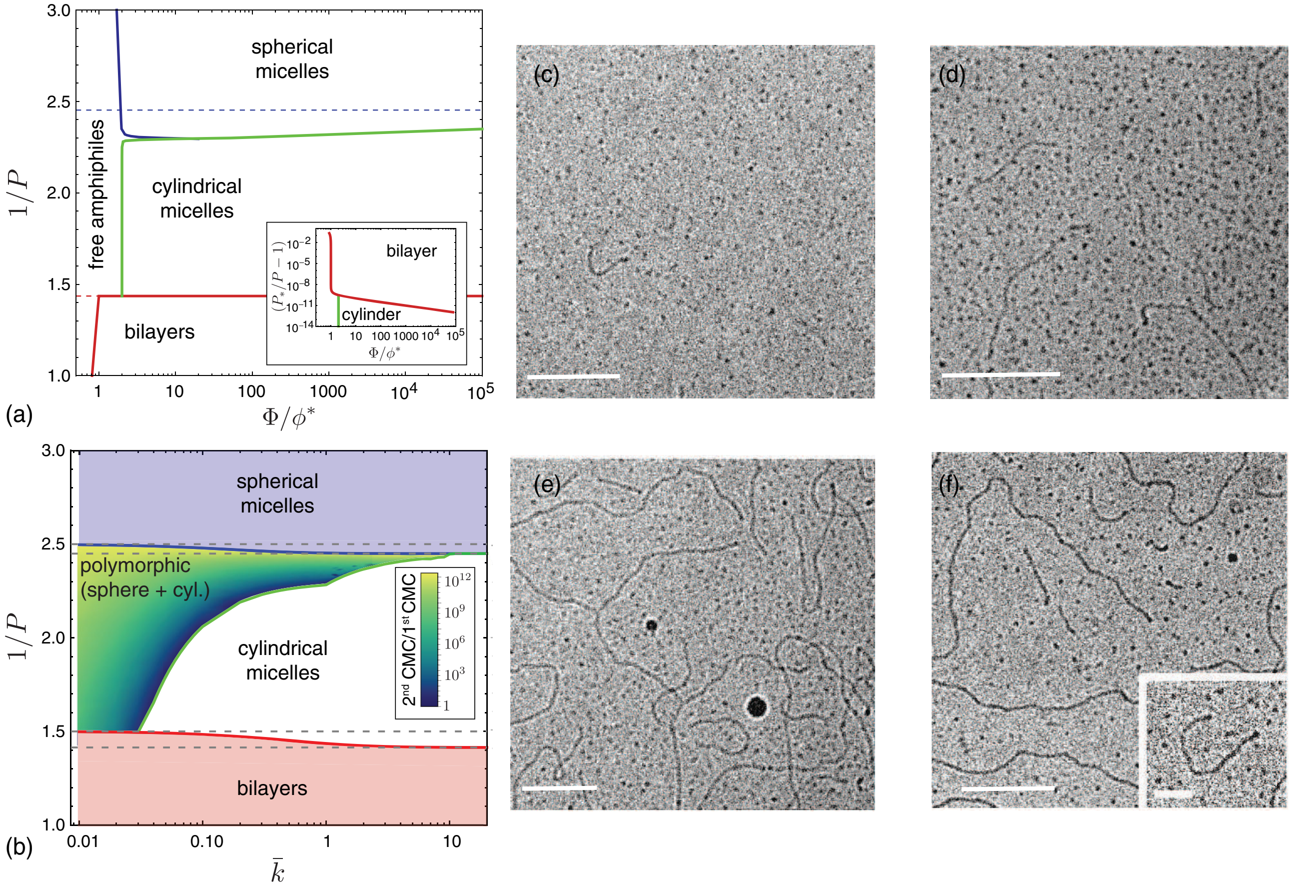

FIG. 12. (a) Assembly state phase diagram for the amphiphilic aggregate model in Eq. (44) for dimensionless stiffness $\bar{k} = 1$ and $v_0^2/a_0^3 = 10$, where ϕ_* is the nominal CMC for cylinders. Solid lines mark the boundaries between the most populous aggregate type, and dashed lines indicate the boundaries in the infinite Φ limit. Inset: enlargement near the boundary between cylinders and bilayers illustrating an extremely narrow window of secondary CMC behavior due to the large mean (finite) size of cylinders. (b) Summary of the polymorphic assembly of the amphiphile model in the plane of stiffness $\bar{k}$ and inverse packing parameter P^{-1} for $v_0^2/a_0^3 = 10$. Regimes of single CMC behavior are shown in solid red, white, and blue for bilayers, cylinders, and spheres, respectively. The regime of polymorphic concentration-driven sphere-to-spherocylinder transitions is colored on a blue-green scale according to the ratio of the second CMC (spheres to spherocylinders) to the first CMC (monomers to spheres). (c)–(f) Cryo-transmission electron micrographs of micelles formed by dimeric (gemini) surfactancts, at 25 °C at increasing weight percent: (c) 0.26%, (d) 0.5%, (e) 0.62%, and (f) 0.74% (scale bars equal 100 nm). The coexistence of cylindrical micelles of increasing total length for (d)–(f), and absence of lengths intermediate to spheres and the shortest cylinders is consistent with a concentration-dependent second CMC. The inset of (f) shows the bulbous ends of a cylindrical micelle (scale bar, 25 nm), which is consistent with an energy barrier between spherical and long cylinders due to a mismatch in preferred radius. (c)–(f) Adapted from Bernheim-Groswasser, Zana, and Talmon, 2000.

(larger radius) sphere. We note that this spherocylinder geometry will overestimate the cost of end caps relative to a more realistic model that includes, say, variation of the radial thickness along the micelle (May and Ben-Shaul, 2001).

As described in Sec. II.B.3, the existence of multiple low-energy branches of aggregation leads to a rich phase behavior, which can be analyzed according to the law of mass action [Eq. (5)]. As summarized in Appendix A, we analyze the polymorphic assembly of the model described by Eq. (44), making several simplifying assumptions. Specifically, we adopt the continuum limit for n, calculate the optimal spherocylinder radius for each length L by minimizing the per-molecule energy [Eq. (44)] as a function of r (i.e., neglecting radius fluctuations), and numerically calculate the mass fractions of spheres, spherocylinders, and membranes as functions of k, P, and Φ from Eq. (5).

Figure 12(a) shows the phase diagram as a function of the packing parameter and total concentration Φ for a dimensionless stretching stiffness $\bar{k} = 1$. Analogous to the phase diagram of the generic model in Fig. 7(b), with increasing packing parameter the system undergoes dimensional transitions from spheres to spherocylindrical assemblies and finally to bulk lamellar aggregates (corresponding to packing dimensionalities of $d_{\mathrm{L}} = 3, 2, 1$). The infinite-concentration limits of the phase boundaries [indicated by dashed lines in Fig. 12(a)] correspond to the packing parameter values where the bulk free energy per subunit of two aggregate geometries are equal: $\epsilon_3 = \epsilon_2$ at the sphere-spherocylinder boundary, and $\epsilon_2 = \epsilon_1$ at the spherocylinder-lamella boundary, with $\epsilon_{d_{\mathrm{L}}}$ the optimal energy per subunit in the aggregate interior for aggregates with dimensionality d_{L}. For values of the packing parameter near each of these dimensional transitions, there is a concentration-dependent dimensional transition. That is, for packing parameter values close to the infinite-concentration value for the spherocylinder-sphere transition $2.45 \gtrsim P^{-1} \gtrsim 2.3$, spheres are favored at low concentrations, with a transition to spherocylinders occurring above a threshold concentration that diverges exponentially as the packing parameter approaches the transition value, i.e., $P^{-1} \to 2.45$. An analogous behavior occurs at the spherocylinder-lamella transition (Fig. 12 inset), but the range of P^{-1} values is exceptionally narrow due to the large (yet finite) mean size of spherocylinder aggregates (see the forthcoming discussion).

In Fig. 12(b), we show an overview of the polymorphic assembly in terms of the two control parameters of the amphiphile aggregation model P^{-1} and $\bar{k}$. Boundaries in the P^{-1}-$\bar{k}$ space are shown for infinite-concentration sphere-spherocylinder and spherocylinder-lamella transitions, indicated by blue and red lines, respectively, with gray dashed lines corresponding to the values of these transitions in the extensionally "floppy" ($\bar{k} \to 0$) and "stiff" ($\bar{k} \to \infty$) limits.[14] We also show the region in P^{-1}-$\bar{k}$ space for which there is a concentration-driven sphere-spherocylinder transition (a secondary CMC), with color indicating the width in concentration space of the transition. More specifically, the color scale indicates the ratio of pseudocritical concentrations $\Phi_{\mathrm{cyl}}/\Phi_{\mathrm{sph}}$ where spherocylinders or spheres, respectively, become the most populous subunit state (i.e., greater than 50%) at a given value of stiffness and packing parameter.

The emergence of the polymorphic, concentration-driven transition between spherical and spherocylindrical micelles captured in Fig. 12 can be understood in terms of the three ingredients of the secondary CAC behavior encoded in Eq. (35): the energy gap between minimal-energy spheres and infinite cylinders $\Delta\epsilon = \epsilon_3 - \epsilon_2$, the finite aggregation number in spherical micelles n_{sph}, and an energy barrier δ separating the spherical aggregates from spherocylinders [as in Fig. 11(d)]. Below the blue curve in Fig. 12(b), where $\Delta\epsilon > 0$, the physical origin of the intermediate concentration state of spheres is the higher (per subunit) entropy associated with their fewer subunits. This window of second CMC behavior widens in Φ as the gap between infinite cylinders and spheres vanishes to zero, which happens as the inverse packing parameter increases and approaches the blue curve in Fig. 12(b). Likewise, from Eq. (44) it is straightforward to see that the gap between $d_{\mathrm{L}} = 3$ and 2 vanishes as $\bar{k} \to 0$ since the is no obstacle to achieving the optimal head-group packing ($a \to a_0$) for any d_{L} in the absence of extensional stiffness. Hence, the ratio of the second CMC (spheres to spherocylinders) relative to the first (monomers to spheres) grows large in both of these regimes.

The regime of second CMC behavior is restricted to lower values of extensional stiffness and disappears above a threshold value of $\bar{k}$ due to its effect on the energy barrier between the spheres and spherocylinder micelles. In the limit of $k \to 0$, the thickness of spheres and cylinders is determined purely by head-group packing, and hence $d_{\mathrm{L}} = 3$ and 2 micelles have different radii, implying a finite frustration cost for the hemispherical end caps of the spherocylindrical micelles. As summarized in Eq. (35), the window of second CMC behavior is widened with increasing energy barrier between compact and 1D assemblies. With increasing $\bar{k}$, this barrier diminishes, ultimately vanishing in the $\bar{k} \to \infty$ limit, because the high stiffness requires the micelle thickness to maintain $\ell = \ell_0$ independent of dimensionality. Hence, as is the case for the ladder model of cylindrical micelle thermodynamics (Missel *et al.*, 1980), in the absence of an energy barrier between spheres and elongated cylinders there is only a single CMC in a state where mean aggregation number continuously increases with Φ.

Similar arguments apply to the spherocylinder-lamella transition, except that spherocylindrical aggregates have a small translational entropy due to their large mean size, and thus are stabilized by entropy over a vanishingly narrow region of parameter space.

Evidence for the secondary CMC transition between spherical and cylindrical micelles has been reported for range of surfactant systems; see Bergström (2016) for a review. Many experiments have reported an indirect signature of an inflection point, and secondary upturn, in the mean aggregation number as a function of concentration. For example, the convex dependence of viscosity on concentration for certain ionic surfactants was interpreted by Porte *et al.* (1984) as a secondary CMC and provided motivation for their model with an energetic gap between spherical and cylindrical states for second CMC behavior. Elsewhere, experimental imaging of the concentration-dependence micelle morphologies has been used to probe second CMC behavior. For example, Figs. 12(c)–12(f) show cryo-EM images of micelles formed by dimeric (gemini) surfactants (Bernheim-Groswasser, Zana, and Talmon, 2000). Above the first CMC, but still below a second threshold concentration, cryo-EM shows monodisperse spherical micelles (Fig. 12). With increased concentration [Figs. 12(d)–12(f)], cryo-EM shows the appearance of cylindrical, or wormlike, micelles. While these grow in length with concentration, they also coexist with a population of spherical micelles. This, along with an observed gap in micelle sizes intermediate to spheres and the shortest cylinders, is consistent with the secondary CMC transition induced by an energetic gap between spheres and cylinders. As previously described, and consistent with the apparently "bulbous" ends of wormlike micelles [Fig. 12(f)], such an energetic gap is a natural consequence of the mismatch between radii of spherical and cylindrical micelle packings.

To conclude the overview of amphiphilic aggregation, we return to the question of the convexity of the energetics described by Eqs. (40)–(44) for spherical micelles ($d_{\mathrm{L}} = 3$). It is straightforward to consider two simple limits to estimate the dependence of convexity on target size. When $\bar{k} \ll 1$, energetics are dominated by head-group area terms and $\epsilon''_{\mathrm{T}}(\bar{k} \ll 1) \propto \gamma v_0^{2/3}/n_{\mathrm{T}}^{7/3}$, whereas the opposite limit is controlled by length elasticity, and $\epsilon''_{\mathrm{T}}(\bar{k} \gg 1) \propto k v_0^{2/3}/n_{\mathrm{T}}^{4/3}$. As with the fluid capsule model in Sec. III.A.1, convexity

---

[14] The transition values between d_{L} and $d_1 + 1$ are $P^{-1} = [d(d-1)]^{1/2}$ and $P^{-1} = d_{\mathrm{L}} - 1/2$ in the stiff ($\bar{k} \to \infty$) and floppy ($\bar{k} \to 0$) limits, respectively.

generically decreases with target size, but not nearly as quickly. In particular, for the amphiphile model the product $n_{\rm T}^3 \epsilon''_{\rm T}/k_B T$ is always *increasing* with $n_{\rm T}$ (i.e., as $n_{\rm T}^{2/3}$ and $n_{\rm T}^{5/3}$ for small and large $\bar{k}$, respectively). Hence, according to Eq. (20) the relative number fluctuations of micelles *decrease* with mean size. This is in marked contrast to the predictions for the bending elasticity of the fluid shell model, for which $\epsilon''_{\rm T} \propto n_{\rm T}^{-3}$ and size fluctuations grow in proportion to mean size. The origin of this relatively sharper minima of aggregation energetics, and the correspondingly tighter control of aggregate size, can be traced to the fact the closure radius is on the scale of the subunit thickness (i.e., $R_{\rm close} \approx t$) or, more specifically, the additional considerations of tail packing constant density in the micellar cores, which are absent for the membrane bending elasticity of fluid capsules.

B. Self-limited, open-boundary assembly

Here we describe a class of self-limiting assembly characterized by an open boundary, a surface that separates the aggregate interior from the solution of freely associating subunits. Unlike the previously described SCA, in open-boundary assemblies (OBAs) this free boundary is maintained in the target of self-limiting states. Therefore, the target state has a finite surface energy associated with loss of short-range cohesion or differences in solvation at its exterior.

As described in Sec. III.A.1, a finite boundary energy alone generically favors unlimited aggregates. Hence, to be self-limited OBA structures require additional interaction terms that grow with aggregate dimensions, and thus balance the generic tendency to minimize the boundary to interior ratio. We define this additional nonsurface energetics as the excess energy, and its essential feature is a regime of superextensive growth, meaning that the total excess energy increases with size faster than the number of subunits n. We later describe two examples of mechanisms that generate such a form of excess energy, but the key underlying feature is the existence of gradients in stress throughout the aggregate. Whereas SCA can realize finite target dimensions with uniform subunit shape and packing, in OBA long-range gradient patterns of intra-aggregate stress are required for self-limitation and ultimately dictate the range of possible self-limiting sizes.

Before introducing these two physical mechanisms, which illustrate the microscopic origins of excess energy accumulation, we begin with a generalized description of the thermodynamics of OBA.

1. Limits of self-limitation

To describe the aggregation energetics of OBA, we consider a structure with an open boundary that can grow in D possible spatial dimensions (directions). We let $d_{\rm L}$ of these directions be potentially self-limiting, while in the remaining $d = D - d_{\rm L}$ directions the structure undergoes unlimited growth. Denoting the limited and unlimited dimensions of the structure as W and L, respectively, we find that the scaling of aggregate volume gives $n \propto W^{d_{\rm L}} L^d$. In the limit that $L \gg W$, the amount of the open boundary then grows as $A_{\rm b} \sim W^{d_{\rm L}-1} L^d$.[15] For example, consider subunits that can bind in all three spatial dimensions, i.e., $D = 3$. Quasicylindrical aggregates of such subunits accrue a surface energy cost derived from their full boundaries, which are limited in two transverse spatial dimensions ($d_{\rm L} = 2$) and unlimited in the axial direction ($d = 1$), whereas finite-thickness planar aggregates are limited along the normal direction ($d_{\rm L} = 1$), but unlimited in the two in-plane directions ($d = 2$). Such structures are self-limiting spatially but do not necessarily have a finite or even well-defined peak aggregation number.[16] Nevertheless, equilibrium in these cases (such as finite-thickness filaments and slabs) derives from the optimal energy per subunit with respect to the self-limiting dimension(s) of the assembly.

Because surface subunits typically have fewer cohesive bonds with neighbor subunits and potentially more unfavorable contacts with surrounding solvents, an open boundary generically accrues a surface energy cost proportional to $A_{\rm b}$. Parametrizing this cost by the boundary energy Σ and the bare aggregation energy for interior subunits $-\epsilon_0$, we may write the following generic form for the per subunit aggregation energy as a function of the self-limiting dimension W:

$$\epsilon(W) = -\epsilon_0 + \frac{\Sigma}{W} + \epsilon_{\rm ex}(W). \tag{46}$$

For simplicity, we absorbed a geometric factor associated with the dimensionality of the boundary into the definition of Σ. The first two terms describe the short-range cohesive interactions, with a constant bulk energy gain and a surface energy penalty, while the final term defines the excess energy relative to the short-range model. An example of energy of this form is shown in Fig. 13(a). As we later illustrate, such an excess energy arises from effects such as long-range inter-subunit repulsions or elastic stresses that increase with aggregate size. In self-limiting assemblies, these cumulative effects give rise to an excess energy per subunit that increases monotonically in size; hence, $\epsilon_{\rm ex}(W)$ captures energetic effects that grow *superextensively* with aggregate size. In any physical system, this superextensive energy growth will only persist up to some threshold assembly size, crossing over from convex (e.g., power-law) growth at small sizes to some asymptotically saturating energy density as $W \to \infty$. The large-W saturation of excess energy can occur for a variety of reasons. For example, long-range repulsions may be screened beyond some length scale. Alternatively, above a threshold excess energy cost, it will become energetically favorable for subunits to reorganize or deform to avoid further excess energy accumulation. Notwithstanding its microscopic origin, the effect of this saturating excess energy is to renormalize the per subunit energetics from its bare value to $-\epsilon_0 + \epsilon_\infty$.

---

[15] For $D = 2$, assemblies are sheetlike and the boundary corresponds to a 1D edge due to fewer lateral cohesive bonds, while for $D = 3$ the boundary corresponds to the entire 2D surface surrounding the aggregate.

[16] Indeed, cylindrical aggregates have an exponential length distribution, while planar aggregates (such as plates and membranes) correspond to a bulk state according to the analysis in Sec. II.A.2.

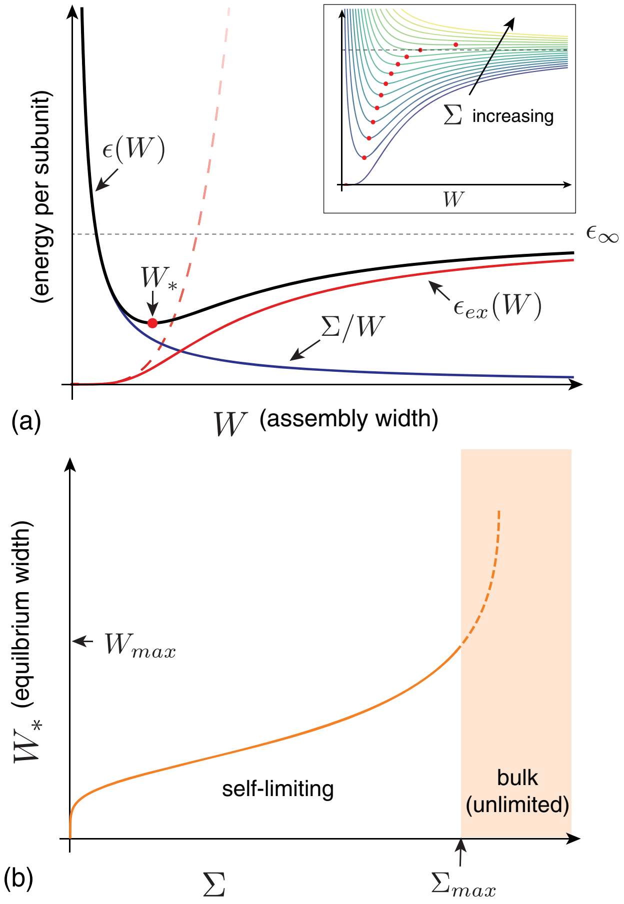

FIG. 13. (a) Schematic plot of the aggregation energetics per subunit for self-limiting open-boundary assembly, where the blue, red, and black curves, respectively, show the surface energy density, excess energy, and total energy density. The dashed curve shows power-law growth of excess energy at small size. The red point highlights the equilibrium self-limiting width W_*. Inset: variation of this equilibrium width with increasing surface energy Σ assuming a fixed $\epsilon_{\rm ex}(W)$. (b) Schematic plot of the equilibrium width as a function of Σ, where the dashed portion indicates the possibility of a finite-width branch that is metastable relative to the bulk state $W \to \infty$. The boundary between equilibrium self-limiting and bulk states is marked by a maximum self-limiting size $W_{\rm max}$, denoted as the escape size.

In OBA, self-limitation follows directly from the balance between the accumulating cost of $\epsilon_{\rm ex}(W)$ and the generic decrease of surface energy with increasing W. It is then straightforward to show that equilibrium assemblies satisfy the following equation of state that links the finite equilibrium size W_* to the surface energy:

$$\Sigma = W_*^2 \epsilon'_{\rm ex}(W_*), \tag{47}$$

with $\epsilon'_{\rm ex} = \partial_W \epsilon_{\rm ex}$. Stability criteria additionally require that $\epsilon''_{\rm ex} > 0$, but the basic results of the competition between surface energy and excess energy accumulation are shown schematically in Fig. 13. Since the surface energy always drives assembly toward larger sizes, the equilibrium finite size W_* generically increases with Σ given a fixed form of $\epsilon_{\rm ex}(W)$. Self-limitation can arise in two ways: either the equilibrium width can increase continuously with Σ to the bulk state (i.e., $W_* \to \infty$) or, as illustrated in Fig. 13(b), self-limitation will persist only up to a maximal finite size before a discontinuous transition to bulk assembly occurs. In the latter case, the energy density of the finite state eventually increases with surface energy, to the point where the energy densities of the finite and bulk unlimited states become equal; i.e., $\epsilon(W \to \infty) = -\epsilon_0 + \epsilon_\infty$. For surface energies above this maximal value, the bulk unlimited state is favored. Hence, such systems can be characterized by a maximal self-limiting size $W_{\rm max}$ and a maximal surface energy $\Sigma_{\rm max}$, below which equilibrium structures are finite [Fig. 13(b)].

For a general OBA, it is then useful to consider the following question: What is the range of possible self-limited equilibrium states that a given system can exhibit? As described in Sec. II.B.2, provided that the concentration is well above the aggregation threshold (CAC), the mean aggregate size is determined by the minimum of the per subunit energy $\epsilon(n)$. Thus, the answer to this generic question depends only on the excess energy and its accumulation with width.[17]

To see this, we reformulate the condition for equilibrium of the self-limiting state relative to bulk assembly $[\epsilon(W_*) < \epsilon(W \to \infty)]$ in terms of the surplus of energy in the bulk relative to the finite state

$$\Delta\epsilon(W_*) \equiv \epsilon(W \to \infty) - \left(\frac{\Sigma}{W_*} + \epsilon_{\rm ex}(W_*) - \epsilon_0\right) \tag{48}$$

$$= \frac{\partial}{\partial W_*}\{W_*[\epsilon_\infty - \epsilon_{\rm ex}(W_*)]\}, \tag{49}$$

where we use the equation of state linking stable size to surface energy in Eq. (47). The condition that $\Delta\epsilon(W_*) > 0$ is required for equilibrium finite self-limited states can then be simply formulated in terms of the first integral of surplus bulk energy

$$\mathcal{A}(W) \equiv W[\epsilon(W \to \infty) - \epsilon(W)] = W[\epsilon_\infty - \epsilon_{\rm ex}(W)] - \Sigma, \tag{50}$$

so the equilibrium of finite structures relative to bulk corresponds to the condition

$$\left(\frac{\partial \mathcal{A}}{\partial W}\right)_\Sigma > 0. \tag{51}$$

We refer to the function $\mathcal{A}(W)$ as the *accumulant* and note that graphically it corresponds to the area of the rectangular regions of the plot of $\epsilon_{\rm ex}(W)$ vs W highlighted in Fig. 14(a) for model 1, which is described by the blue curve. Because the fixed-Σ partial derivative in Eq. (51) is independent of surface energy, for a given form of excess energy the accumulant may be constructed for any value of surface

---

[17] Equation (19) shows a small concentration-dependent shift of the optimal size n^* below the size corresponding to the minimum of $\epsilon(n)$; the optimal size approaches the energy-minimizing size as $\Phi \to \infty$.

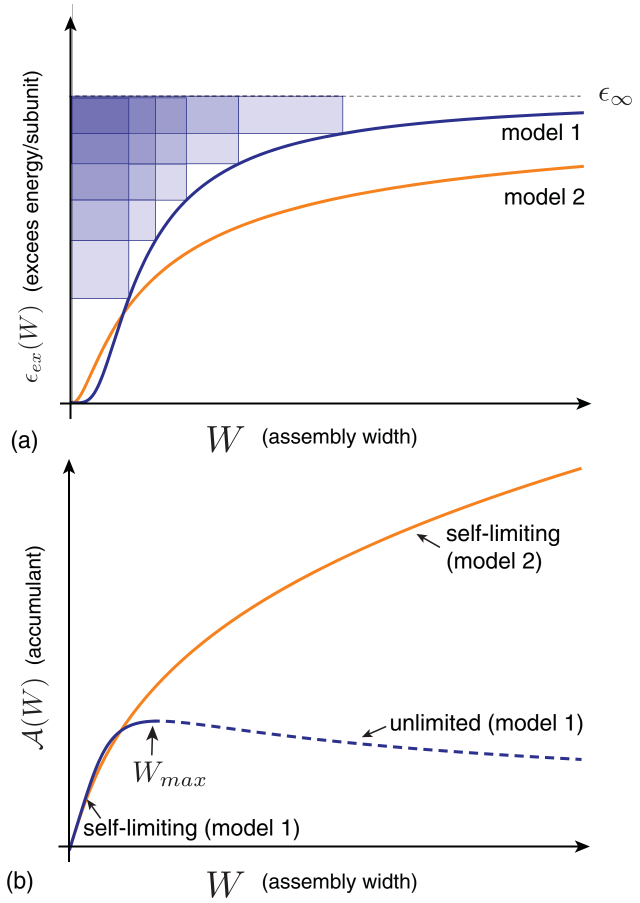

FIG. 14. (a) Schematic plots of excess energy vs width for two different models of OBA: model 1 (blue curve) and model 2 (orange curve). Both cross over from power-law growth at small size to saturating excess energy at large size but do so with different functional forms (the asymptotic approach to ϵ_∞ is slower in model 2 than in model 1). The area of the blue rectangles graphically illustrates the definition of the accumulant in Eq. (50). (b) Accumulant $\mathcal{A}(W)$ as a function of width for the two models shown in (a). Regions of increasing $\mathcal{A}(W)$ correspond to possible ranges of equilibrium self-limiting sizes for a given form of $\epsilon_{\mathrm{ex}}(W)$. That is, values of W for which $\mathcal{A}'(W) > 0$ correspond to energy minima of $\epsilon(W)$, which are lower in energy than the bulk state $\epsilon(W \to \infty)$, for a particular value of the surface energy Σ given by Eq. (47). Hence, model 1 shows an upper limit for maximal self-limiting size W_{max}, while model 2 exhibits stable self-limiting equilibria at all sizes.

energy and analyzed as a function of W to consider potential finite-size equilibria at all values of Σ.

According to Eq. (51), finite-W equilibria correspond to the range of increasing $\mathcal{A}(W)$. Figure 14 illustrates two models of excess energy, both of which are characterized by crossovers from power-law growth at small W to asymptotic saturation to finite values of $\epsilon_{\mathrm{ex}}(W \to \infty) = \epsilon_\infty$ but do so via different functional dependencies on the finite size W.

For model 1 (blue curve), the range of increasing $\mathcal{A}(W)$ extends only up to a maximum, with corresponding width W_{max}, indicating that a first-order transition between finite and bulk states occurs at W_{max} and the corresponding value of Σ_{max}.[18] For model 2 (orange curve), the monotonically increasing range of the accumulant extends to $W \to \infty$. This indicates that model 2 supports equilibrium self-limited states at all values of Σ and reaches bulk assembly only in the limit of infinite surface cost. Hence, even if the excess energy saturates in the $W \to \infty$ limit, self-limitation may still, in principle, extend to all possible size ranges, depending on the nature of the asymptotic approach to the bulk energy.

Given the form of the accumulant defined in Eq. (50) and the condition (51) for its increase, the possibility for self-limited states that extend continuously up to the bulk state (i.e., $\lim_{\Sigma \to \infty} W_* \to \infty$) can be deduced from the asymptotic form of the residual energy $\Delta\epsilon(W) = \epsilon_{\mathrm{ex}}(\infty) - \epsilon_{\mathrm{ex}}(W)$ as $W \to \infty$. Following an argument made by LeRoy (2018) and Terzi, LeRoy, and Lenz (2020), we assume that this residual vanishes as a power law $\epsilon_{\mathrm{ex}}(\infty) - \epsilon_{\mathrm{ex}}(W) \sim W^{-\nu}$. It is then straightforward to show that when $\nu > 1$, $\mathcal{A}(W)$ decreases as $W \to \infty$. Such cases correspond to the first-order type self-limitation exhibited by model 1 in Fig. 14. Alternatively, when $0 < \nu < 1$, indicating a slower saturation of excess energy, the accumulant continues to increase as $W \to \infty$, indicating the existence of self-limited equilibria extending up to the bulk state. The case of $\nu = 1$ is marginal and can exhibit either first- or second-order-like behavior. Later we describe models in the context of geometrically frustrated assemblies that can illustrate both types of behavior, involving either a *continuous* or *discontinuous* transition between the finite and bulk states depending on the mechanisms underlying the accumulation of excess energy.

2. Short-range attractions, long-range repulsions

As described in Sec. III.B.1, self-limitation in OBA requires superextensive growth of the excess energy. In this first class of examples, the accumulation of excess energy derives from long-range interactions between subunits, specifically, interactions characterized by short-range attraction and long-range repulsion (SALR) (Groenewold and Kegel, 2001; Sciortino *et al.*, 2004). Models of such systems usually consider isotropic pair potentials $u(r)$ that can be split into two parts,

$$u(r) = u_{\mathrm{SA}}(r) + u_{\mathrm{LR}}(r), \tag{52}$$

where the short-range potential describes cohesive interactions $u_{\mathrm{SA}}(r) < 0$ between neighboring subunits, which act on scales comparable to the subunit hard-core diameter d, i.e., $u_{\mathrm{SA}}(r \gg d) \simeq 0$. Outside of this cohesive range, the potential is dominated by a long-range repulsion $u_{\mathrm{LR}}(r) > 0$ that extends over sizes much larger than single particles. Even when the repulsive interactions between neighboring subunits are much weaker than the cohesion, the fact that repulsive interactions extend far beyond the first shell of neighbors can lead to superextensive growth of repulsive energy with increasing assembly size. A model of this form has been applied to explain finite-sized aggregate formation in a broad

---

[18] Defining the accumulant in terms of the $\Sigma = 0$ energetics, i.e., $\mathcal{A}(W) \equiv W[\epsilon_\infty - \epsilon_{\mathrm{ex}}(W)]$, as in Fig. 14(b), it can easily be shown that $\Sigma_{\mathrm{max}} = \mathcal{A}(W_{\mathrm{max}})$.

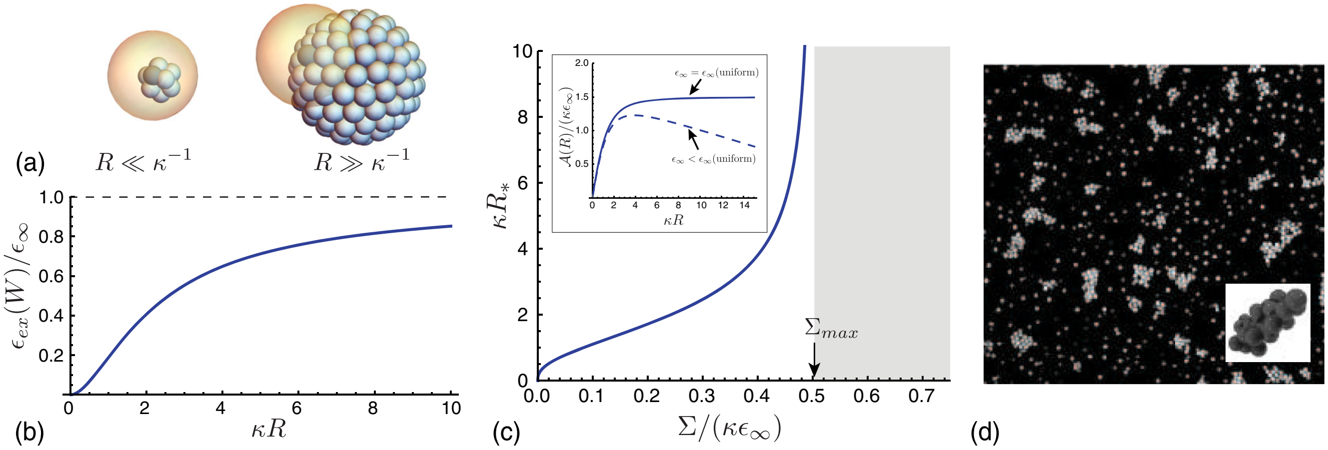

FIG. 15. (a) Schematics of aggregates of particles interacting via short-range (i.e., contact) interactions and long-range repulsions, with a screening length κ^{-1}, which can be large or small relative to the aggregate size R. (b) Excess energy of spherical aggregates vs radius showing a crossover from power-law growth for $R \ll \kappa^{-1}$ to an asymptotically saturating bulk energy for $R \gg \kappa^{-1}$. (c) Plot of the equilibrium finite radius R_* of the SALR model vs surface energy Σ showing a continuous divergence to the bulk at a finite $\Sigma = \Sigma_{\text{max}}$. The second-order-like transition from finite to bulk states predicted by this model (solid curve) is consistent with the monotonically increasing form of the accumulant plotted in the inset, where this result assumes that the bulk state has uniform density. Assuming instead that the bulk state has a lower-energy, nonuniform density [such as a periodic aggregate morphology (Zhang *et al.*, 2019)], the transition will be first order, as illustrated by the nonmonotonic accumulant shown as a dashed line. (d) Confocal microscope image of clusters of charged colloidal particles (radius 660 nm). In these experiments short-range attractions are induced by depletion forces generated by inert polymers in suspension, while electrostatic repulsions are maintained at relative long range due to the low concentration of mobile ions in the host organic solvent. Adapted from Stradner *et al.*, 2004.

range of systems (Dinsmore, Dubin, and Grason, 2011), including protein complexes (Stradner *et al.*, 2004; Cardinaux *et al.*, 2007; Foderà *et al.*, 2013), charged nanoparticles (Nguyen *et al.*, 2015), colloidal particles in low-dielectric solvents (Sedgwick, Egelhaaf, and Poon, 2004; van Schooneveld *et al.*, 2009), dipolar mesophases (Seul and Andelman, 1995), and even models of nuclear matter (Caplan and Horowitz, 2017).

To illustrate the mechanism of self-limitation in this class of systems, we consider the following specific model: Short-range cohesive interactions lead to $-u_0$ per neighbor contact, and aggregates maintain an approximately uniform density $\rho_0 \approx d^{-3}$. For repulsive interactions, we assume that subunits are isotropic and repel according to a screened Yukawa repulsion

$$u_{\text{LR}}(r) = \frac{q^2}{r} e^{-\kappa r}. \tag{53}$$

Here q is the electrostatic charge per subunit, assumed to be fixed, and κ is the screening length, which arises from Debye-Hückel screening by mobile ions in solution and truncates the far-field repulsions for $r \gg \kappa^{-1}$. For simplicity, we illustrate a simplified version of the theory by Groenowold and Kegel (GK) (Groenewold and Kegel, 2001). In our presentation, we consider the case where charge per subunit is fixed, a point that we revisit later. Our primary purpose is to illuminate a model with the minimal ingredients for a self-limiting equilibrium and, due to its finite screening length, the long-range Yukawa potential provides a convenient example. For purposes of illustration, we consider spherical aggregates with radius R, whose interaction free energy can be described by

$$\epsilon(R) = -u_0 \frac{\langle z \rangle}{2} + \frac{3\Sigma}{R} + \epsilon_{\text{ex}}(R), \tag{54}$$

where $\langle z \rangle$ is the mean number of neighbor contacts in the bulk of the aggregate and $\Sigma \approx u_0/d^2$ is the surface energy associated with fewer short-range cohesive contacts at the boundary. Here the excess energy derives directly from the sum of long-range, pairwise repulsions in the aggregate volume $\mathcal{V}$ as[19]

$$\begin{aligned} \epsilon_{\text{ex}}(R) &= \frac{q^2\rho_0}{2\mathcal{V}} \int_{\mathcal{V}} d^3\mathbf{r}\, d^3\mathbf{r}'\, u_{\text{LR}}(|\mathbf{r} - \mathbf{r}'|) \\ &= \frac{q^2\rho_0}{2\kappa^2} \left\{ 1 - \frac{3(1+\kappa R)}{2(\kappa R)^3} (1 + e^{-2\kappa R})[\kappa R - \tanh(\kappa R)] \right\}. \end{aligned} \tag{55}$$

The behavior encoded in $\epsilon_{\text{ex}}(R)$ is plotted in Fig. 15(b) and can be understood physically by considering the asymptotic limits of small and large aggregate size relative to the screening length of the repulsive interactions, as shown schematically in Fig. 15(a). For small aggregates,

---

[19] The form of the self-energy can be readily calculated using Green's theorem, where the total repulsive energy can be written as $\int_{\mathcal{V}} d^3\mathbf{r}\, q\rho(\mathbf{r})\phi(\mathbf{r})$ with a potential that satisfies the linearized Debye-Hückel equation $(\nabla^2 + \kappa^2)\phi(\mathbf{r}) = 4\pi q\rho(\mathbf{r})$. Solving this for spherically symmetric aggregates $\rho(r \leq R) = \rho_0$ and $\rho(r > R)$ yields the explicit form of screened electrostatic energy.

$$\epsilon_{\mathrm{ex}}(R \ll \kappa^{-1}) \simeq \frac{q^2 \rho_0}{5} R^2, \tag{56}$$

which derives from the fact that, when repulsions extend over the entire aggregate, the per subunit cost of the Coulomb self-energy of an aggregate is roughly $Q^2/(nR) \propto R^2$, where the total aggregate charge $Q = qn \propto R^3$. In the opposite regime, where aggregate sizes far exceed the screening length,

$$\epsilon_{\mathrm{ex}}(R \gg \kappa^{-1}) \simeq \epsilon_\infty \left(1 - \frac{3}{2\kappa R} + \frac{3}{2(\kappa R)^3}\right), \tag{57}$$

where $\epsilon_\infty \equiv q^2 \rho_0 \kappa^{-2}/2$. This leading term $R \to \infty$ derives from the fact that each subunit in the bulk of the aggregate experiences repulsive interactions with roughly $\rho_0 \kappa^{-3}$ other subunits within a screening length, while the first correction accounts for the surface layer of thickness κ^{-1} with fewer neighbors within the screening length. The subleading $1/R^3$ term can be associated with a square-curvature cost per unit area for deforming the boundary shape from planar, which alters the distribution of repulsive particles near the free surface of the aggregate.

Figure 15(c) shows the predicted equation of state for the self-limiting radius R_* as a function of surface energy $\Sigma(R_*) = R_*^2 \epsilon'_{\mathrm{ex}}(R_*)/3$. For small Σ, the balance between the charging energy of aggregates and surface energy leads to a growth $R_* \sim \Sigma^{1/3}$, which proceeds until the optimal aggregate size grows beyond the screening length. In the large aggregate regime, the asymptotic approach to ϵ_∞ leads to an aggregate size that diverges continuously at a critical surface energy $\Sigma_{\mathrm{max}} = \kappa \epsilon_\infty/2$. The origin of the second-order-like transition to the bulk in this model can be traced to the $R \gg \kappa^{-1}$ form of the excess energy in Eq. (57). The leading correction to $\epsilon_{\mathrm{ex}}(R \to \infty) = \epsilon_\infty$ goes as $-1/R$, that is, in the notation of Sec. III.B.1, $\nu = 1$. Hence, this leading correction at large R behaves like a negative contribution to the surface energy $\Sigma_{\mathrm{eff}} = \Sigma - \Sigma_{\mathrm{max}}$ due to the reduced electrostatic repulsion within a screening length of the surface. When $\Sigma < \Sigma_{\mathrm{max}}$, the effective surface energy is *negative* and the aggregate equilibrium derives from the balance between the $-1/R$ drive to create more surface and the subleading $+1/R^3$ term to give

$$R_* \sim \sqrt{\frac{\epsilon_\infty \kappa^{-3}}{\Sigma_{\mathrm{max}} - \Sigma}}, \tag{58}$$

which diverges continuously as $\Sigma \to \Sigma_{\mathrm{max}}$. This prediction, that there is a maximal cohesive surface energy for aggregates but that their equilibrium self-limiting size extends to arbitrarily large values, is consistent with the plot in the Fig. 15(c) inset, which shows the accumulant of spherical aggregates to be monotonically increasing over the full range of R.

The result that self-limitation can extend up to arbitrarily large sizes (i.e., $W_{\mathrm{max}} \to \infty$) for this simplified model may be surprising, as it implies that the thermodynamics are sensitive to the finite size of aggregates over much larger size ranges than the finite interaction range κ^{-1}. That is, a subunit whose interactions extend only κ^{-1} can still "sense" that it should join an aggregate with a radius smaller than $R_* \gg \kappa^{-1}$, but not a larger one. The resolution of this puzzle is that the physical term that restrains aggregate growth in this $R_* \gg \kappa^{-1}$ regime derives from the square-curvature cost of the free boundary (contributing as $+1/R^3$ to the excess energy). Although repulsions are short ranged compared to large aggregates (i.e., $R \gg \kappa^{-1}$), they are sufficiently nonlocal to sense the *curvature* of the boundary and, thereby, the global radius of the aggregate. That is, repulsions in this regime give rise to precisely the type of square-curvature energetics that selects for finite sizes in self-closing assemblies in Sec. III.A.1, i.e., with a preferred boundary curvature that vanishes as $\Sigma \to \Sigma_{\mathrm{max}}$.[20]

While self-limited assembly due to a competition between cohesive boundary costs and accumulation of long-range repulsion is generic, the specific features of this simplified model, including continuously diverging finite size for $R_* \gg \kappa^{-1}$, require several caveats. Foremost, the convexity of the assembly energetics decreases with aggregate size as $\epsilon''_* \sim n_*^{-3}$ for $R_* \gg \kappa^{-1}$ for this model, similar to the case of self-closing shells in Sec. III.A.1. Thus, according to Eq. (20), fluctuations in n grow with self-limiting size and diverge at the threshold as $\langle (\Delta n)^2 \rangle^{1/2} \propto n_* \sim (\Sigma_c - \Sigma)^{-3/2}$. Hence, for all practical purposes the self-limitation to finite and well-defined sizes will not persist to arbitrarily large structures.

Beyond this, the model is oversimplified and fails to incorporate a number of physical ingredients, which can influence the form of excess energy accumulation. For example, more realistic models would include a finite compressibility of the short-range cohesive forces, which allows for density variation with aggregate size and position within the aggregate and, for charged subunits in solvent, the effects of dielectric contrast between the aggregates and the solvent, as well as variable degrees of charge condensation and dissociation as aggregates vary in size. Indeed, when initial applications of this GK model aimed to understand the aggregation of lysozyme proteins in low-salt aqueous solutions, comparison to scattering measurements of the aggregate size suggested that variable ionization of subunits could not be neglected (Zaccarelli, 2007). While to a first approximation it was argued that subunit charging is independent of aggregation number (Groenewold and Kegel, 2001), the charge per subunit and the screening length, which varies with counterion concentration, both vary with total concentration of ionizable subunit species (Cardinaux *et al.*, 2007). Hence, accounting quantitatively for the concentration dependence of aggregation for these electrostatic systems requires consideration of subunit charge in a self-consistent fashion, leading potentially to aggregation-dependent renormalization of repulsive interactions (Nguyen *et al.*, 2015), not to mention additional effects associated with nonspherical aggregate shapes (Sciortino, Tartaglia, and Zaccarelli, 2005). The nonspherical nature of aggregates of charged particles is evident in confocal images of a colloidal analog to the aggregation of nanoscopic proteins (Sedgwick, Egelhaaf, and Poon, 2004), as shown Fig. 15(d).

---

[20] Consistent with this result, performing an analogous calculation for a planar slab geometry shows that the self-limited size diverges exponentially above a threshold value of Σ due to the absence of such a curvature term.

Finally, even beyond specific physical considerations of electrostatic SALR systems, the prediction of the simplified Yukawa model of diverging finite aggregate size assumes a spatially uniform bulk state, whereas models of systems with short-range attractions and long-range repulsions have been shown to form periodically modulated aggregate phases, such as stripes, layers, and spheres, at high concentration (Seul and Andelman, 1995; Sear and Gelbart, 1999; Sciortino *et al.*, 2004; Zhuang and Charbonneau, 2016). That is, at sufficiently high densities, the uniform density bulk state is unstable to lower free energy periodic bulk states such that $\epsilon_\infty(\text{periodic}) = \epsilon_\infty(\text{uniform})$. If the stability of equilibrium self-limitation is reanalyzed in terms of a competition with a lower-energy, nonuniform bulk state, then the maximal size range of self-limiting equilibrium becomes finite; see the accumulant plot in Fig. 15(d). The full thermodynamics of the transition between a dilute phase of self-limiting aggregates (above the CAC) to the long-range ordered bulk aggregate phases requires consideration beyond the ideal aggregation thermodynamics considered here.

3. Geometrically frustrated assembly (GFA)

The notion of geometric frustration (GF) originally emerged in the context of low-temperature condensed matter systems (magnetic materials, spin models, etc.) (Vannimenus and Toulouse, 1977). It refers to the impossibility of propagating an energetically preferred arrangement throughout space due to global geometric constraints (Kléman, 1989; Sadoc and Mosseri, 2006). For bulk, infinite systems, GF leads to a rich phenomenology: extensive arrays of topological defects thread through the highly degenerate bulk ground states that populate a rough energy landscape.

Recently it has been recognized that GF gives rise to new behaviors in self-assembling materials (Grason, 2016) derived from two key features. First, the constituent subunits (polymers, colloids, proteins, etc.) are relatively "soft" and held together by weak, noncovalent forces. Second, assemblies need not reach bulk states and thus have additional degrees of freedom associated with the potentially finite size and shape of the assembled domain. Unlike bulk or rigid systems where GF must be resolved by defects (Sadoc and Mosseri, 2006), in soft assemblies it can be tolerated, at least over some size range, by smooth gradients in the subunit shapes and packings. As an illustration, see the schematic of "warped jigsaw" particles in Fig. 16(a), where the tapering of the particle shape favors curvature along one row of the lattice assembly (Grason, 2017). Provided that particles or their interactions are sufficiently deformable, aggregates can accommodate frustration through strain gradients, leading to arrangements that are more (less) relaxed near (far from) the free boundary of the aggregate.

This self-organization of long-range stress gradients is the defining characteristic of GFA, giving rise to the form of accumulating excess energy that can limit assembly size (Sec. III.B.1). The balance between the surface energy and the superextensive cost of GF can select equilibrium domain sizes that are finite and, in principle, arbitrarily larger than the subunits themselves.

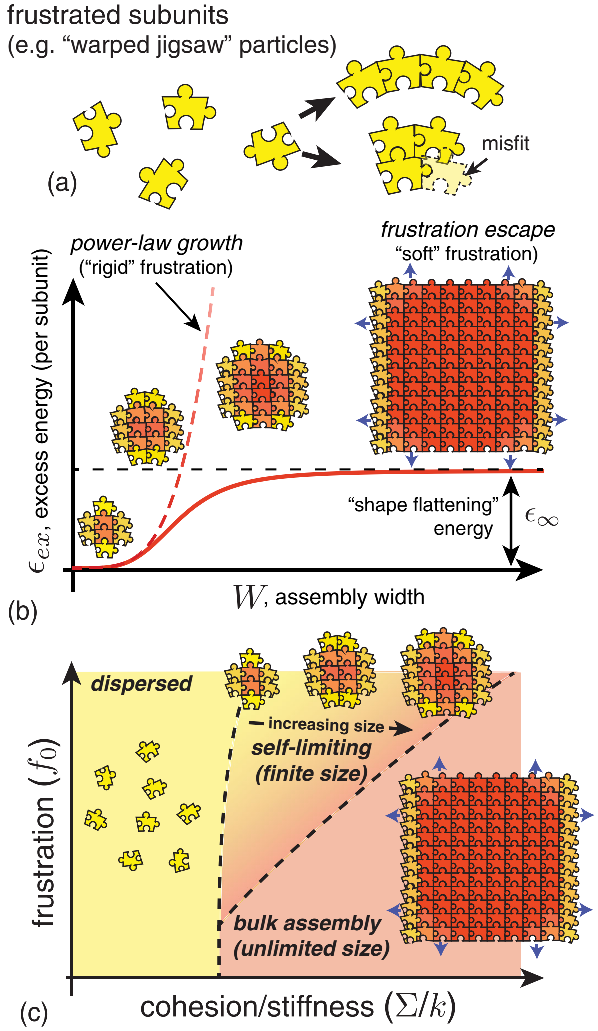

FIG. 16. (a) Heuristic warped jigsaw model for GFA from Grason (2017), in which directional interactions promote curvature along rows of a locally preferred 2D "lattice." Assembly in both bonding directions leads to "misfits," as in the tetrameric aggregate. (b) Schematic plot of the excess energy per subunit for the warped jigsaw model due to the superextensive buildup of elastic costs of misfits. Intra-assembly strains are illustrated via particle color, from unstrained (yellow) to highly strained (red). The excess energy shows the characteristic crossover from power-law growth at small aggregates to an asymptotic approach to a strained bulk state (in this case envisioned as shape flattening of jigsaw particles). (c) Schematic phase diagram for a generic model GFA in Eq. (59), considered (at fixed concentration and temperature) as a function of the ratio of surface energy to elastic stiffness and a measure of the strength of frustration f_0.

To date GFA has been implicated in the emergent structures of various soft matter systems, including self-twisting protein bundles (Aggeli *et al.*, 2001; Turner *et al.*, 2003; Grason and Bruinsma, 2007; Yang, Meyer, and Hagan, 2010; Brown, Kreplak, and Rutenberg, 2014; Hall *et al.*, 2016; Cameron, Kreplak, and Rutenberg, 2018), twisted molecular crystals (Haddad *et al.*, 2019; Li *et al.*, 2020), chiral smectics (Hough *et al.*, 2009; Matsumoto, Alexander, and Kamien, 2009) and membranes (Selinger *et al.*, 2004; Ghafouri and Bruinsma, 2005; Gibaud *et al.*, 2012; Armon *et al.*, 2014; Sharma *et al.*, 2014; Kang and Lubensky, 2017; Sakhardande *et al.*, 2017), particle-coated droplets (Bausch, 2003; Irvine, Vitelli, and Chaikin, 2010; Meng *et al.*, 2014; Yu, Ghosh, and Hagan, 2016), curved protein shells (Zandi *et al.*, 2004; Li *et al.*, 2018), and phase-separated lipid vesicles (Schneider and Gompper, 2005). Consideration of GF in these systems has primarily stemmed from experimental observations of assemblies that (a) terminate at finite size and/or (b) exhibit defect-ordered morphologies. Continuum models that consider the interplay between the elastic costs of misfit and domain formation have been developed to address a range of distinct frustration mechanisms, including frustration of 2D liquid-crystalline or crystalline order on non-Euclidean manifolds (Nelson and Peliti, 1987; Bowick and Giomi, 2009), metric and orientational frustration of chiral fibers (Brown, Kreplak, and Rutenberg, 2014; Grason, 2015; Haddad *et al.*, 2019), shape frustration in stacking assemblies of curved layers (DiDonna and Kamien, 2003; Achard *et al.*, 2005; Matsumoto, Alexander, and Kamien, 2009), chirality frustration in crystalline (Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014) and liquid-crystalline membranes (Selinger, Spector, and Schnur, 2001; Kang and Lubensky, 2017), and the assembly of nontiling polygonal particles (Lenz and Witten, 2017).

We first describe a heuristic model that highlights the common thermodynamic features of these apparently diverse realizations of GFA, which are encoded in a simplified form of the excess energy

$$\epsilon_{\mathrm{ex}}(W) \approx \frac{k}{2}\delta^2(f, W) + \frac{C}{2}(s - s_0)^2. \tag{59}$$

Here k is a generalized elastic parameter for straining interelement packing, as measured by a generalized mean strain $\delta(f, W)$, which itself varies with the finite size W of the domain and a parameter f that measures the strength of frustration. To be clear, local strains vary with position throughout the assembly, as we later describe for a specific example, but for simplicity we focus here on how the magnitude of strain varies with size and frustration. While frustration mechanisms vary considerably among distinct GFA systems, they all exhibit power-law growth of strain with domain size W. This can be modeled heuristically as

$$\delta \approx f(s) W^{\eta}, \tag{60}$$

where η and f vary for different cases of GFA. For example, the orientational strains in 2D liquid crystal domains grow linearly with domain size ($\eta = 1$), whereas positional strains in frustrated 2D crystals grow quadratically ($\eta = 2$) (Grason, 2016; Niv and Efrati, 2018). The definition of frustration strength f depends on the specific GF mechanism; however, it can generally be expressed as a function $f(s)$ of the local shape of inter-subunit packing (such as inter-subunit bend or twist), which we denote generically with the shape parameter s. In many cases, the frustration strength can be expressed as a simple power law of shape

$$f(s) \approx s^{\mu}, \tag{61}$$

where μ is a positive exponent. For the example of a crystalline cap on a spherical surface (Grason, 2016), f corresponds to the Gaussian curvature, which is the square of the 1D curvature in this geometry and thus corresponds to a shape parameter with $\mu = 2$. In the form of Eq. (61), the strength of frustration generically vanishes, intuitively, in the limit in which the shape flattens to $s \to 0$. The second term in Eq. (59) describes generic costs for deformations away from an ideal, misfitting shape with $s = s_0 \neq 0$, which incur elastic penalties described by the shape modulus C.

This basic form of Eq. (59) implies a generic size dependence for the excess energy shown schematically in Fig. 16(b). For small sizes, assemblies retain their preferred, misfitting shape ($s \simeq s_0$) leading to a power-law growth of excess energy $\epsilon_{\mathrm{ex}}(W \to 0) \simeq k f_0^2 W^{2\eta}/2$, where $f_0 = f(s_0) = s_0^{\mu}$ is the frustration strength of the preferred shape. If power-law growth of $\epsilon_{\mathrm{ex}}(W)$ is extended to all size scales, then the compromise between costs of GF and surface energy would select a finite equilibrium size $W_* \sim f_0^{-2/(2\eta-1)}$ for any surface energy Σ. However, in any physical system, the excess energy can accumulate only up to some maximal size scale, beyond which the assemblies *escape frustration* through one of a number of competing morphological "modes" (Hall *et al.*, 2016; Hall and Grason, 2017). These include, for example, the formation of topological defects that screen far-field frustration stresses (Grason, 2012; Li *et al.*, 2019), as well as shape flattening, which refers to the smooth deformation of an incompatible (i.e., misfit) shape to a uniformly strained, compatible one (Grason, 2020). Because subunits and their interactions are generically soft, the excess energy required to escape frustration must be finite. Thus, as shown in Fig. 16(b), $\epsilon_{\mathrm{ex}}(W)$ generically crosses over from power-law accumulation at small W to an asymptotic approach to this finite energy $\epsilon_{\mathrm{ex}}(W \to \infty) = \epsilon_{\infty}$. In the minimal description of Eq. (59), the cost of shape flattening is simply $\epsilon_{\infty} = C s_0^2/2$ per subunit.

Comparing power-law accumulation at small sizes to the asymptotic shape-flattened energy, one expects a crossover between these regimes at a characteristic flattening size $W_{\mathrm{flat}} \approx [(C/k)/s_0^{2(\mu-1)}]^{1/\eta}$. This size scale defines the maximum size for which assemblies will tolerate the accumulating frustration cost, beyond which deformation to an unfrustrated shape ($s \to 0$) becomes energetically favorable. Intuitively, this length scale also sets a bound on the escape size $W_{\mathrm{max}} \leq W_{\mathrm{flat}}$ since for $W \gg W_{\mathrm{flat}}$ the bulk energy is simply renormalized by the cost of flattening. In general, the escape size is set by the lowest-energy mode of relaxing frustration in the bulk state, which may also involve Wigner lattice states of defects that neutralize the long-range cost of frustration (Li *et al.*, 2019). Understanding the practical limit of self-limitation then requires one to consider all possible competing modes of relaxing frustration and determine which of these has the lowest energy for a particular regime of assembly. This distinction between a power-law growth of energy density with W and an asymptotic approach to a constant value was proposed by Meiri and Efrati (2021) to classify distinct regimes of GF itself as, respectively, cumulative and noncumulative frustration. The analysis of the accumulant in Sec. III.B.1 translates that criteria directly into their thermodynamics consequences for equilibrium SLA.

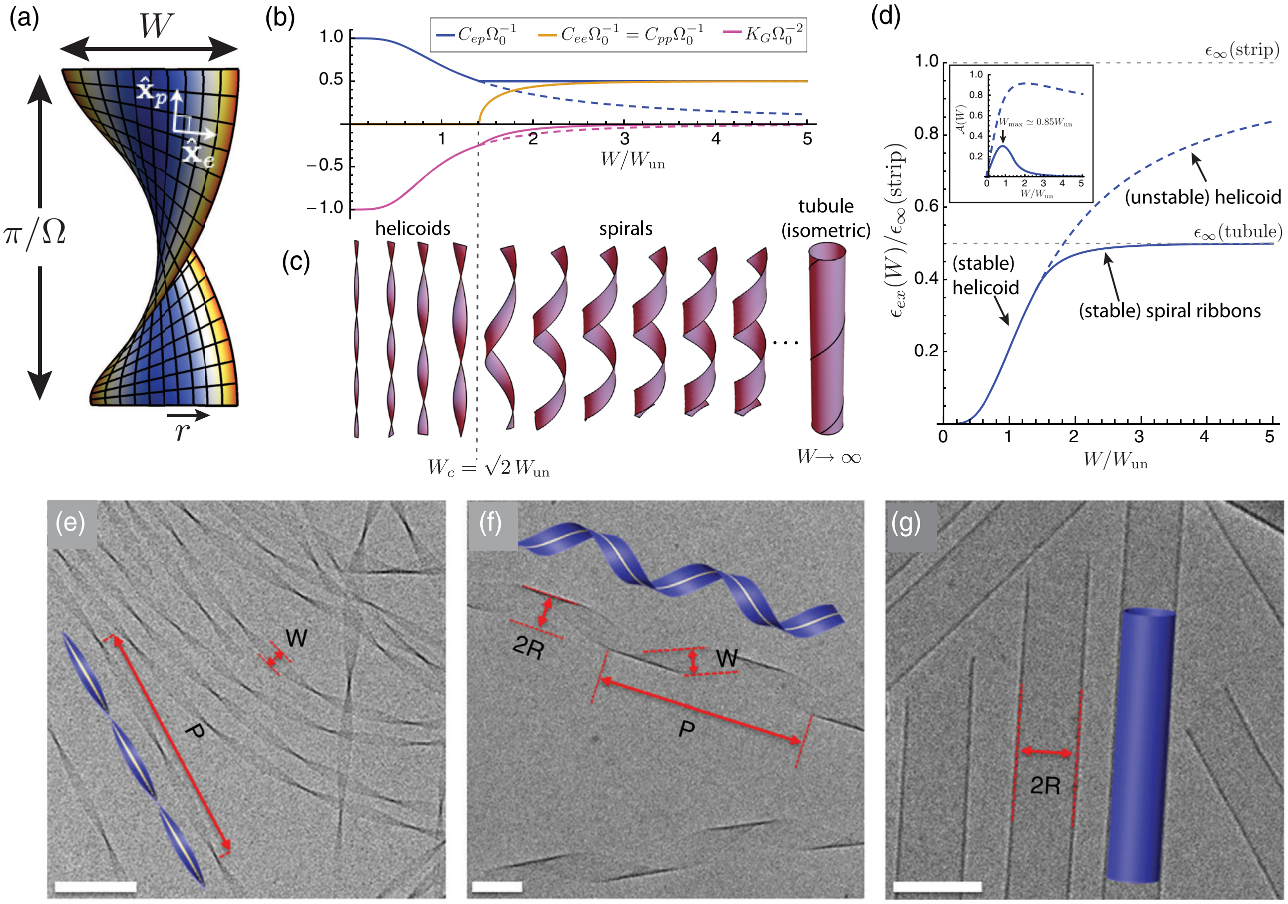

FIG. 17. (a) Schematic of a chiral, crystalline ribbon, where colors indicate the local extensional strains required by negative Gaussian curvature, from low (blue) to high (red). (b) Plots of the equilibrium shape relaxation of the narrow-ribbon model of Ghafouri and Bruinsma (2005) as a function of increasing ribbon width for curvature components C_{ij} and Gaussian curvature $K_{\rm G}$, which is approximately uniform over the ribbon. The solid branches show the minimal-energy configurations, while the dashed line indicates the unstable helicoidal equilibrium. Widths are scaled by the characteristic length $W_{\rm un}$ defined in Eq. (63). (c) Schematics of the shape equilibrium, in particular, the shape transition from helicoids to spiral ribbons at a critical width $W_{\rm c}$. (d),(e) Respective excess energies for helicoidal (dashed lines) and helicoid-spiral shape branches (solid lines). While both shape modes predict an asymptotically relaxed frustration energy and a finite self-limiting width, spiral ribbons expel $K_{\rm G}$ at a much faster rate, leading to a narrower range of self-limitation. (e)–(g) Transmission electron microscopy images of ribbons assembled from chiral bola-amphiphiles. Illustrated are morphologies evolving with assembly time (scale bars, 100 nm): in (e) helicoids are observed after 24 hours, in (f) spiral ribbons are observed after one week, and, finally, in (g) closed tubules are observed after 5 months. Adapted from Zhang *et al.*, 2019.

Notwithstanding which mode facilitates escape to the bulk, the heuristic picture of GFA implies a common phase diagram [Fig. 16(c)] spanned by the bare frustration strength f_0 on one axis and the ratio of cohesion to intra-assembly stiffness Σ/k on the other (say, for fixed subunit concentration and temperature). Above a critical frustration strength, there is a regime of self-limiting aggregates between the dispersed state (below the CAC) and the bulk state. Within the self-limited regime, the equilibrium domain size increases with Σ/k but decreases with frustration strength f_0.

As a specific example of the origin and implications of GFA, we now discuss a well-studied model for crystalline ribbons frustrated by chirality (Selinger *et al.*, 2004; Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014). This model has been developed to understand the polymorphic assembly of chiral surfactant bilayers that adopt a variety of quasi-1D structures (Oda *et al.*, 1999; Selinger, Spector, and Schnur, 2001; Ziserman *et al.*, 2011; Zhang *et al.*, 2019); i.e., they form 2D sheets that are much narrower in width than length ($W \ll L$). Ghafouri and Bruinsma (2005) first described this mechanism of frustration, posing it as a competition of the bending cost of a chiral anisotropic membrane against the elastic costs of in-plane stretching for a 2D crystal with Gaussian curvature. More recently this model has been extended (Armon *et al.*, 2014; Grossman, Sharon, and Diamant, 2016) to describe experimental observations of bola-amphiphile ribbons by Zhang *et al.* (2019).

For this discussion, we present here only a simplified picture of the frustration and its effects in chiral ribbons and provide a more detailed summary in Appendix B. As shown in Fig. 17(a), we consider assemblies of ribbons of width W and much longer (unlimited) length $L \gg W$. For the case of sufficiently narrow ribbons, it can be shown that the favored morphology is that of a *helicoid*: a strip whose width axis twists around a straight central ribbon at a uniform rate Ω. In this regime, the excess energy of the helicoidal chiral ribbon takes the following form (see Appendix B):

$$\epsilon_{\mathrm{ex}}(W) = \frac{Y a_0}{1440} \Omega_0^4 W^4 + B a_0 (\Omega - \Omega_0)^2, \tag{62}$$

where a_0 is the area per subunit in the membrane and Y and B are the respective in-plane stretching and out-of-plane bending moduli for the assembled membrane. Here Ω plays the role of a shape parameter in the previously introduced heuristic model, and the second term has the form of a chiral bending energy that favors uniform twist Ω_0 (Helfrich and Prost, 1988; Ghafouri and Bruinsma, 2005). The first term derives from the elastic cost of in-plane deformation of the 2D lattice of the ribbon and describes the accumulating elastic cost of frustration. It is intuitive to understand the growth of stretching energy with ribbon width by considering the contour length per helical turn of the longitudinal strip of the ribbon at a distance r from the center $(2\pi/\Omega)\sqrt{1+(\Omega r)^2}$. Because of the resistance to shearing and stretching of local elements in the 2D solid, this motif leads to strains that grow with relative longitudinal extension compared to the midline, $\approx(\Omega r)^2$ to lowest order, and hence it generates an elastic energy density of $\sim Y(\Omega^2 W^2)^2$. It can be shown that the underlying source for stress gradients in such an assembly derives from the incompatibility of a 2D planar metric with the nonzero negative Gaussian curvature of the membrane $K_{\mathrm{G}} \simeq -\Omega^2$ (Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014). In the language of the previously introduced heuristic picture, we can identify the strength of frustration with K_{G}, and hence $\eta = \mu = 2$, due to the quartic growth of frustration cost with shape misfit Ω and width W.

An instructive, albeit oversimplified, analysis of the size dependence of frustration in chiral ribbons considers only the relaxation of helicoidal twist for variable width. For small widths, the ribbon adopts a preferred twist $\Omega_*(W \to 0) = \Omega_0$ due to the vanishing cost of frustration as $W \to 0$ and $\epsilon_{\mathrm{ex}} \approx Y\Omega_0^4 W^4$. For large widths $W \to \infty$, the balance between stretching and bending favors unwinding of the pitch, $\Omega_* \simeq \Omega_0 (W_{\mathrm{un}}/W)^{4/3}$, where

$$W_{\mathrm{un}} \equiv \left(\frac{720B}{Y\Omega_0^2}\right)^{1/4} \tag{63}$$

defines a characteristic unwinding size. That is, for $W \ll W_{\mathrm{un}}$ the ribbon retains roughly the twist preferred by chirality and the power-law ($\sim W^4$) accumulation in excess energy, while for much larger ribbons the prohibitive cost of frustration causes the ribbon to unwind, expelling Gaussian curvature. Figure 17(d) (dashed curve) shows the characteristic crossover in the excess energy of helicoidal ribbons, with an asymptotic flattening of the excess energy characterized by the exponent $\nu = 4/3$, which implies that self-limiting widths are not stable in the unwinding region; see Sec. III.B.1. In terms of the accumulant analysis [Fig. 17(d) inset], the escape of frustration by helicoid unwinding would imply a maximum self-limiting size of $W_{\max}(\text{helicoid}) \simeq 2.1 W_{\mathrm{un}}$.

How do we understand physical parameters that determine the range of possible self-limiting widths? Here we note that it arises from the combination of two physical lengths. One of these derives from the ratio of bending to stretching moduli $\sqrt{B/Y} \equiv t$, which is typically of the order of the thickness of an elastic membrane. In other words, it is expected that t is of the order of the molecular ($\approx$ nanometer) size of the constituent amphiphiles. The second length scale is the preferred pitch $P_0 = 2\pi/|\Omega_0|$, which derives from the chiral preference for local skew packing of neighbor amphiphiles (Zhang *et al.*, 2019). Unlike t, the size range of P_0 is mesoscopic, of the order of hundreds of nanometers. With these definitions, we see that the unwinding size scale, and hence the maximum self-limiting width, is the geometric mean of these two length scales $W_{\mathrm{un}} \simeq 11\sqrt{tP_0}$, one molecular and one mesoscopic. It is then understood that frustration can limit the size of helicoids up to a size range intermediate to these molecular and mesoscopic sizes, that is, ribbons of the order of tens of nanometers in width.

As we describe in more detail in Appendix B, frustration escape in chiral ribbons is more complex than suggested by considering only helicoidal shapes. As chiral ribbons grow beyond a critical width $W_{\mathrm{c}} = \sqrt{2} W_{\mathrm{un}}$ [Figs. 17(b) and 17(c)], they become mechanically unstable to a new class of shape equilibria, *spiral ribbons* (Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014). With increasing width, this class of shapes approaches an isometric strip that is wound helically around a cylinder. Hence, unlike the helicoid, this more complex shape relaxation allows the assembly to retain some residual chiral twist while at the same time expelling the Gaussian curvature that includes in-plane stresses. Therefore, these spiral shapes facilitate a more efficient escape of frustration than helicoidal unwinding; see Fig. 17(d).

Notwithstanding the quantitative frustration relaxation for large widths, it can be shown that the mechanical instability does not change the basic conclusion, namely, that the scale of the self-limiting width is set by the geometric mean of the t and P_0. That is, according to the accumulant analysis of the spiral ribbon branch [Fig. 17(d) inset], $W_{\max} \simeq 0.85 W_{\mathrm{un}}$, and the instability only reduces the numerical prefactor of the unwinding scale. This is consistent with a recent experimental study of ribbon morphologies of bola-amphiphiles (Zhang *et al.*, 2019) [Figs. 17(e)–17(g)]. In this system, the molecular size suggests that $t \approx 3$–4 nm, while skewed packing of chiral neighbors in the crystal leads to much larger pitches of 200 nm. The observation of helicoidal ribbons up to $\approx$50 nm (consistent with the geometric mean of molecular size and helical pitch) demonstrates that frustration stress can propagate far beyond molecular dimensions. When ribbons grow beyond this size, they do not grow to infinitely wide and untwisted sheets but instead transition to a second mechanism of self-limitation, forming closed tubules of finite diameter proportional to the pitch.

The preceding discussion has neglected relaxation of frustration by defects. For 2D solid chiral ribbons this can be justified because the critical width for defect formation in helicoids[21] far exceeds the transition to a shape-flattening

---

[21] This can be conjectured to be $\propto |\Omega_0|^{-1}$ based on standard arguments of Gaussian curvature screening; see Bowick and Giomi (2009).

state, whose lower energy is facilitated by the soft bending modes of thin solids. This conclusion does not hold in general, even when the same morphologies arise, if the underlying mechanism of frustration differs. For example, chiral ribbons with only liquid-crystalline (LC) in-plane order (e.g., hexatic) would be described by the same shape (bending) elasticity, but with a frustration cost that arises from orientational strains from Gaussian curvature (Vitelli and Turner, 2004; Mbanga, Grason, and Santangelo, 2012). Such angular strains grow with a weaker power law $\sim K_{\mathrm{G}} W$ than positional strain (Niv and Efrati, 2018), and hence are an example of $\eta = 1$ strain growth. This "softer" growth of frustration energetics implies a shape-flattening transition that takes place at a much larger size that is proportional to the mesoscopic pitch $|\Omega_0|^{-1}$. This is a size scale at which disclinations may also be expected to lower the ribbon energy, implying that frustration escape for LC ribbons likely falls into a different class than for solid ribbons, one that may mix both smooth (shape-flattening) and singular (defect-mediated) modes.

IV. KINETIC PATHWAYS TOWARD SELF-LIMITING EQUILIBRIUM

As the primary focus of this review is the equilibrium ingredients and thermodynamics of SLA, we have not considered the nonequilibrium pathways by which such systems, starting from an out-of-equilibrium initial condition, may arrive at a self-limited equilibrium distribution. However, the kinetics of assembly can have significant influence over the size distributions and resulting morphologies formed in experimental systems of SLA systems since they are necessarily limited to observations at finite times. Hence, interpretation of practically all experiments must allow for the possibility of nonequilibrium effects. In this section, we provide a basic introduction to a few of the key principles for understanding and modeling kinetics of SLA. Rather than giving an exhaustive review, our purpose is largely to illustrate how the language and formalism introduced in the foregoing discussions of equilibrium SLA translates to the basic conceptual and theoretical frameworks used for analyzing these same systems out of equilibrium. Far more thorough and general descriptions of the interplay between kinetics and assembly products, in both bulk and SLA processes, were given in several previous reviews (Sear, 2007; Agarwal and Peters, 2014; Hagan, 2014; Whitelam and Jack, 2015).

Theory, computation, and experiments have shown that the ability of a self-assembling system to approach equilibrium within practical timescales depends on a delicate balance between thermodynamic and kinetic effects (Zlotnick *et al.*, 1999; Ceres and Zlotnick, 2002a, 2002b; Endres and Zlotnick, 2002; Whitesides and Grzybowski, 2002; Zlotnick, 2003; Hagan and Chandler, 2006; Jack, Hagan, and Chandler, 2007; Nguyen, Reddy, and Brooks, 2007; Wilber *et al.*, 2007; Rapaport, 2008; Whitelam *et al.*, 2009; Wilber *et al.*, 2009; Grant, Jack, and Whitelam, 2011; Hagan, Elrad, and Jack, 2011; Klotsa and Jack, 2011; Cheng, Aggarwal, and Stevens, 2012; Grant and Jack, 2012; Hagan, 2014; Whitelam and Jack, 2015). In a broad variety of systems, yields are nonmonotonic with the strength of cohesive interactions that drive assembly. Optimal assembly occurs when the cohesive free energy is on the order of $(5\text{–}10)k_BT$ per subunit-subunit contact depending on the initial subunit concentration and valency of the subunit contacts. To understand this behavior, we require a consideration of the kinetics of assembly.

To begin, we consider a typical bulk *in vitro* assembly kinetics experiment, such as solution assembly of reconstituted proteins into multiunit structures. To prepare an initial state, the system is equilibrated under conditions that do not lead to assembly; i.e., the total subunit concentration is below the CAC ($\Phi < \phi^*$) and thus the equilibrium aggregate distribution is peaked at monomers (see Fig. 4). In practice, this is accomplished by setting a low subunit concentration, a solution pH and salt concentration that ensure weak subunit-subunit attractions, or sufficiently high temperature that the translational entropy of unassembled subunits dominates over inter-subunit attractions. The system is then rapidly quenched into a condition that favors assembly ($\Phi > \phi^*$), either by increasing the subunit concentration Φ or by decreasing the CAC above the current concentration, by changing temperature or physicochemical changes that increase subunit-subunit interactions (such as changes to pH and salt concentration). That is, the new equilibrium state (postquench) corresponds to a population of aggregates coexisting with monomers as described in Sec. II. However, the prequench distribution is out of equilibrium, consisting predominantly of unassembled monomers, and as such gradients in the system free energy will drive the assembly toward lower free energy states with assembled clusters. Note that this scenario can describe experiments of both unlimited (e.g., bulk crystals) and self-limited (e.g., capsules) assembly equally well.

The study of assembly kinetics is concerned with the timescale required to approach the equilibrium state, as well as the structures and lifetimes of long-lived metastable states that may occur along the way. To understand the influence of kinetics on practical applications or biological functions, these timescales must be compared to those that are accessible to an experimenter or a biological organism.

A. Classical nucleation theory and assembly timescales

A useful starting point for understanding the dependence of assembly timescales on control parameters is given by the framework of classical nucleation theory; see Becker and Döring (1935), Binder and Stauffer (1976), Oxtoby (1992), De Yoreo and Vekilov (2003), Hagan and Elrad (2010), Agarwal and Peters (2014), and Hagan (2014), as well as Oxtoby (1992) for a review. In this approach, the assembly of a cluster is broken into two phases: nucleation and growth (often called "elongation" in the context of finite structures). Nucleation refers to the process of overcoming a free energy barrier to form a small but relatively stable aggregate, while the second phase describes growth of such an aggregate to its final optimal size. We see next that for most assembly reactions to be productive (i.e., observable on experimentally realistic timescales) nucleation must be the rate-limiting process, and thus the assembly timescale can be estimated by calculating the nucleation rate.

1. Nucleation kinetics

Consider a general form of aggregation free energy for aggregates that describes assembly driven by short-range cohesive interactions, with possible additional terms to describe higher-order effects such as those that give rise to SLA:

$$\epsilon(n) = -\epsilon_{\mathrm{min}} + \frac{\Delta_0}{n^{1/d}} + \epsilon_{\mathrm{ex}}(n), \tag{64}$$

where ϵ_{min} is the bulk energy per subunit in the aggregate with optimal size, Δ_0 accounts for the surface energy [as in Eq. (6)] and, following Sec. III.B.1, we define $\epsilon_{\mathrm{ex}}(n)$ as the excess energy relative to the bulk and surface effects, including effects that are superextensive in n.[22] Note that this description may apply equally to either self-limiting (i.e., with a minimum at a finite $n = n_{\mathrm{T}}$ and $\epsilon_{\mathrm{T}} = -\epsilon_{\mathrm{min}}$ from Sec. II.B) or unlimited [i.e., where $\epsilon(n)$ is minimal for $n \to \infty$ and $\epsilon_{\mathrm{T}} = -\epsilon_{\mathrm{min}}$ per Sec. II.A.2] assembly. We begin the discussion by considering nucleation for the simplest case $\epsilon_{\mathrm{ex}}(n) = 0$, i.e., unlimited assembly. We see shortly that this analysis also qualitatively applies to most SLA examples since nucleation occurs at small sizes, where the size dependence of the excess energy $\epsilon_{\mathrm{ex}}(n)$ is much smaller than that of the surface terms, whose large values at small sizes constitute the generic origin of the nucleation barrier.

As noted in Sec. II.A.2, small aggregates generically have a smaller cohesive energy than large aggregates because the fraction of subunits at the aggregate surface with unsatisfied interactions [accounted for by the second term Δ_0 in Eq. (6)] decreases with aggregate size. In contrast, the bulk energy ϵ_{min} dominates over surface terms for large aggregates. We define the critical nucleus size n_{nuc} as the crossover between these two regimes. Below n_{nuc} disassembly is favored over assembly because the bulk cohesive energy driving assembly is outcompeted by this unfavorable surface energy and the greater translational entropy of unassembled monomers. Above the critical nucleus size, on the other hand, the bulk cohesion dominates and assembly is favored.[23]

Because forward assembly of prenuclei aggregates is unfavorable, growth of an aggregate to the critical nucleus size is improbable and nucleation is a rare event. In particular, we see that productive assembly requires nucleation to be a rare event on the timescale of typical subunit-subunit association. This gives rise to a separation of timescales: prenucleated aggregates rapidly reach a quasiequilibrium on timescales much shorter than the overall timescale required for the assembly process to approach equilibrium. Thus, based on this assumption, the aggregation distribution for sizes below the critical nucleus size n_{nuc} can be modeled by a variant of the law of mass action

$$\phi_n = n(\phi_1 e^{-\beta\epsilon(n)})^n = n e^{-\beta\Omega(n)} \quad \text{for } n < n_{\mathrm{nuc}}, \tag{65}$$

where $\Omega(n) = n[\epsilon(n) - \mu]$ with $\mu = -k_B T \ln \phi_1$ as the size-dependent grand free energy of n-mers (sometimes referred to as the excess free energy) that accounts for the interaggregate interactions as well as the entropy cost incurred by subunits joining an aggregate. It is important to point out that this nonequilibrium description is a departure from the thermodynamic one introduced in Sec. II, in which μ and ϕ_1 are thermodynamically defined by the total concentration and temperature. Here the chemical potential is used to define only the partial equilibrium of prenuclei with free monomers. Hence, in this usage ϕ_1, and therefore μ and Ω should be understood as time-dependent quantities according to the depletion of free monomers as assembly proceeds. However, the quasiequilibrium approximation assumes that these quantities vary slowly relative to the timescale required for the prenuclei aggregate distribution to reach this form.

By substituting Eq. (6) for the aggregate energy into Eq. (65), we see that (for $d > 1$) there will be a maximum in the grand free energy $\Omega(n)$ at a size

$$n_{\mathrm{nuc}} = \left(\frac{d-1}{d}\frac{\Delta_0}{-\epsilon_{\mathrm{min}} - \mu}\right)^d,$$

owing to the competing drives of negative bulk assembly and positive surface growth. This corresponds to the critical nucleus size since aggregate growth beyond n_{nuc} will decrease the free energy. Here the factor $\Delta_0/(-\epsilon_{\mathrm{min}} - \mu)$ gives a ratio of the unfavorable surface energy Δ_0 that impedes assembly to the net thermodynamic driving force for the assembly $-\epsilon_{\mathrm{min}} - \mu$. This provides a natural measure for how far out of equilibrium the initial conditions are,

$$\Delta\mu \equiv -\epsilon_{\mathrm{min}} - \mu = k_B T \ln\left(\frac{\Phi}{\Phi_{\mathrm{s}}}\right) \geq 0, \tag{66}$$

where we have taken $\phi_1 \simeq \Phi$ since we are considering the initial conditions where nearly all subunits are free. We have defined $\Phi_{\mathrm{s}} \equiv e^{\beta\epsilon_{\mathrm{min}}}$ so that the ratio Φ/Φ_{s} approximately measures how far the total subunit concentration exceeds the CAC and is often referred to as the *supersaturation*.

The key argument of classical nucleation theory is that, because aggregates of the critical nucleus size are rarefied, the nucleation timescale grows exponentially with the barrier height

$$\tau_{\mathrm{nuc}} \sim e^{-\beta\Omega(n_{\mathrm{nuc}})}. \tag{67}$$

There are several important points to make here. First, this is the initial nucleation timescale at the inception of the assembly; the nucleation timescale increases as assembly proceeds because free subunits are depleted and thus μ decreases, in turn increasing $\Omega(n_{\mathrm{nuc}})$. Second, Eq. (65) applies only up to n_{nuc}. Above the critical nucleus size the decreasing free energy implies that growth is relatively rapid, and thus postnuclei aggregates do not reach a quasiequilibrium. Instead, there is a predominant flux of monomers (via association to intermediates) from the population of prenuclei

---

[22] Here the notion of excess energy is shifted by an unimportant constant relative to Eq. (46).

[23] Specifically, critical nucleus denotes a structure for which either complete disassembly or growth to a large aggregate is equally probable. In general, for a particular system there will be an ensemble of critical nuclei that have different structures and (if size is not a complete reaction coordinate) different sizes; see Pan and Chandler (2004).

toward larger aggregates, either toward system-sized aggregates in the case of unlimited assembly or toward a population of target-sized structures in SLA. Third, if the barrier height becomes too small [e.g., $\Omega(n_{\mathrm{nuc}}) \lesssim 10k_BT$ depending on the relevant timescales], the separation of timescales that enabled the quasiequilibrium approximation breaks down, and Eq. (67) will underpredict nucleation timescales since free subunits are rapidly depleted. Fourth, for SLA, additional corrections to this classical nucleation picture will arise (such as an apparent size-dependent surface tension) if the critical nucleus size approaches the finite system size due to the higher-order effects captured by the excess energy (Alder and Wainwright, 1962; Mayer and Wood, 1965; Thompson *et al.*, 1984; Reguera *et al.*, 2003). Fifth, this analysis has assumed that aggregate size n is a good "reaction coordinate," meaning that it accounts for all relevant slow degrees of freedom, and thus the dynamics and probability of an aggregate successfully nucleating can be determined as a function of n. In practice, a complete reaction coordinate must include other aggregate characteristics such as its surface area (Pan and Chandler, 2004). Sixth, a variety of other extensions to classical nucleation theory have been investigated, but it remains highly challenging to quantitatively predict nucleation rates (McGraw and Laaksonen, 1997; Auer and Frenkel, 2001; Peters, 2009; Jacobson, Hujo, and Molinero, 2010; Knott *et al.*, 2012; Loeffler *et al.*, 2012; Joswiak *et al.*, 2013; Statt, Virnau, and Binder, 2015; Zimmermann *et al.*, 2015).

While the discussion thus far has considered an aggregate energy form that drives unlimited assembly, in general, the results are qualitatively similar for SLA with large target structures since the interaction terms that eventually limit assembly grow superextensively and thus are small for small aggregates. As an example, here and in the remainder of this section, we consider the fluid capsid model of Sec. III.A.1 for which self-closing leads to a finite assembly size n_{T}. To simplify the presentation, we assume the limit of high bending modulus $B \to \infty$, so the curvature radius of the assembling shell is fixed to $R_{\mathrm{T}} = \sqrt{a_0 n_{\mathrm{T}}/4\pi}$, and the free energy becomes

$$\epsilon(n) = -\epsilon_{\min} + \tilde{\lambda}\sqrt{\frac{n_{\mathrm{T}} - n}{n}}, \tag{68}$$

with the effective line tension associated with the boundaries of incomplete shells (which gives rise to the nucleation barrier) given by $\tilde{\lambda} = \sqrt{4\pi a_0/n_{\mathrm{T}}}\lambda$, with λ the bare line energy. Notice that, in the limit of small n, Eq. (68) reduces to a bulk cohesive energy $\epsilon_{\min}$ and a surface term $\propto \sqrt{n}$, as anticipated.

This specific $B \to \infty$ model was considered in the context of viral capsid assembly by Zandi *et al.* (2006) and Hagan and Elrad (2010). The grand free energy that corresponds to Eq. (68) is given by

$$\Omega(n) = \Delta\mu n + \tilde{\lambda}\sqrt{n(n_{\mathrm{T}} - n)}, \tag{69}$$

which corresponds to a barrier height (Zandi *et al.*, 2006) of

$$\Omega(n_{\mathrm{nuc}}) = \frac{n_{\mathrm{T}}\tilde{\lambda}}{2}\left(\sqrt{\Gamma^2 + 1} - \Gamma\right), \tag{70}$$

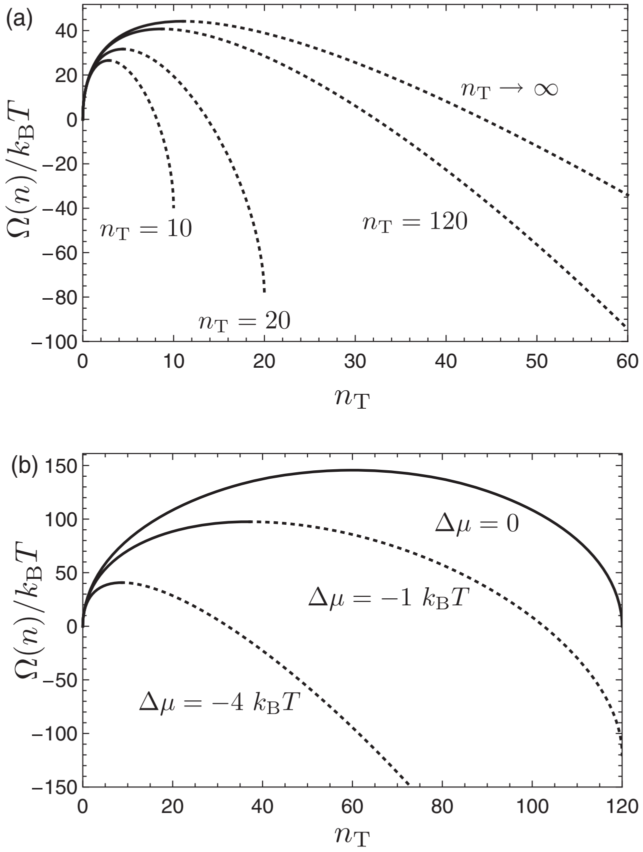

FIG. 18. The grand free energy Ω for the capsid model with fixed curvature R_{T} [Eq. (70)] as a function of partial capsid size n. (a) $\Omega(n)$ for indicated complete capsid sizes n_{T}, with the limit $n_{\mathrm{T}} \to \infty$ corresponding to unlimited assembly (a flat disk). The calculation is performed for a chemical potential difference $\Delta\mu = -\epsilon_{\mathrm{T}} - \mu = -4k_BT$ and a per subunit binding free energy $\epsilon_{\mathrm{T}} = -15k_BT$, corresponding to a subunit-subunit contact energy of $7.5k_BT$ and tetravalent subunits (Zlotnick *et al.*, 2000; Ceres and Zlotnick, 2002b). We set the line energy to $\lambda = \epsilon_{\mathrm{T}}/2a_0^{1/2}$, corresponding to one unsatisfied contact per subunit on the partial shell rim. (b) $\Omega(n)$ as a function of the chemical potential difference $\Delta\mu$ for $\epsilon_{\mathrm{T}} = -15k_BT$ and complete capsid size $n_{\mathrm{T}} = 120$. In (a) and (b), the solid lines correspond to sizes for which the quasiequilibrium approximation described in the text applies (i.e., $n \leq n_{\mathrm{nuc}}$), and thus the concentration of intermediates is given by $\phi(n) \propto e^{-\beta\Omega(n)}$. The dashed lines correspond to sizes $n > n_{\mathrm{nuc}}$ for which this assumption is not valid and the intermediate concentrations cannot be described by a quasiequilibrium (except approximately for the case $\Delta\mu = 0$).

where $\Gamma = \Delta\mu/\tilde{\lambda}$ defines a measure of the dimensionless quench depth,; i.e., the driving force for assembly in the initial state compared to the line energy that impedes the assembly (Zandi *et al.*, 2006).

Figure 18(a) shows the grand free energy as a function of aggregate size for several values of the target shell size n_{T}, as well as for unlimited assembly into a flat disk $n_{\mathrm{T}} \to \infty$. Here we have plotted the portions of the free energy profiles below (above) the critical nucleus size as solid (dashed) lines to emphasize that the quasiequilibrium assumption applies only below the critical nucleus. That is, $\phi(n) \propto e^{-\beta\Omega(n)}$ for $n \leq n_{\mathrm{nuc}}$, but for $n > n_{\mathrm{nuc}}$ there is a constant flux of subunits toward complete shells and the nonequilibrium distribution of intermediate concentrations does not follow the law of mass action.

Notice that the free energy profiles for the self-limited cases are qualitatively similar to that of the unlimited case at small aggregate sizes, which is consistent with the previous analysis. However, the shell geometry that gives rise to SLA does have quantitative effects: the shell curvature causes the length of the unfavorable free boundary to decrease relative to a flat disk (i.e., of equal area). Thus, the critical nucleus size and barrier height increase with target size, asymptotically approaching the flat disk limit. Note that, while in this case the self-closing physics giving rise to SLA decreases the nucleation barrier relative to the unlimited case, SLA effects can in general shift it in either direction. For example, the strain energy that gives rise to GFA in Sec. III.B will increase the barrier beyond the surface costs. We reiterate, though, that the effects of SLA on nucleation are typically quantitative rather than qualitative.

Figure 18(b) shows the grand free energy as a function of aggregate size for several values of the supersaturation, at fixed target size and line tension. We see that as the supersaturation decreases the critical nucleus size and corresponding barrier height increase since the translational entropy of free monomers increases, and thus the net assembly driving force decreases. Note that the different curves corresponding to different supersaturation levels in Fig. 18(a) can be viewed from two perspectives. On the one hand, each supersaturation level can correspond to the initial condition of a separate experiment with a different total subunit concentration Φ, with each curve corresponding to the aggregate free energy at the beginning of the experiment. In this interpretation, Eq. (67) describes the initial nucleation rate and its dependence on quench conditions. On the other hand, as a single experiment proceeds the supersaturation is continually decreasing as free monomers are depleted by assembly, and curves at decreasing supersaturation levels reveal the instantaneous nucleation rate as the experiment proceeds (and subunits are effectively removed from the prenuclei pool). Notice, then, that the barrier height increases as the reaction proceeds, and thus the nucleation rate decreases over time. For large target sizes ($n_\mathrm{T} \gg 1$), as the reaction proceeds the nucleation barrier eventually becomes so large in comparison to the thermal energy that assembly ceases on relevant timescales; thus, the reaction only asymptotically approaches equilibrium.

Finally, the curve corresponding to no supersaturation ($\Delta\mu = 0$) corresponds to the equilibrium state with coexistence of shells and free monomers (hence the entire curve is plotted with a solid line). In this case the critical nucleus size is given by $n_\mathrm{nuc} = n_\mathrm{T}/2$, corresponding to a half shell. Notice that at equilibrium intermediates are higher in free energy than free monomers or complete shells due to the unfavorable line energy at the boundaries and are thus present only at low concentrations; see also Sec. II.B.2.

2. Growth

We define growth as the process by which a critical nucleus assembles to its final state, commensurate with the optimal aggregation size. In contrast to unlimited assembly (such as bulk crystal growth), for SLA there is a well-defined mean timescale for growth of the aggregate since about $n_\mathrm{T} - n_\mathrm{nuc} \approx n_\mathrm{T}$ subunits must associate to reach the optimal size.[24]

The physical process of growth differs fundamentally from nucleation. Nucleation is a highly cooperative process: from Eq. (70), we see that the nucleation timescale decreases exponentially with subunit interaction strength ϵ_min and decreases with subunit concentration according to $\phi_1^{n_\mathrm{nuc}}$. The latter condition reflects the fact that n_nuc subunits must come together within a short timescale in order to create a stable nucleus. In contrast, because postnucleus intermediates are relatively stable, the growth phase can proceed through independent additions of individual subunits or small oligomers. Thus, one can generally expect the growth timescale τ_grow to depend only weakly on ϵ_min. Moreover, one can expect τ_grow to vary inversely with subunit concentration $\tau_\mathrm{grow} \propto \phi_1^{-1}$ since the rate of addition should increase in proportion to concentration unless growth requires overcoming any secondary nucleation barriers. The latter prediction was tested and confirmed experimentally for the hepatitis B capsid assembly by Selzer, Katen, and Zlotnick (2014).

In general, the mean growth timescale can be estimated as

$$\tau_\mathrm{grow} = \frac{n_\mathrm{T}^{\alpha}}{\phi_1 f_\mathrm{assem}}, \tag{71}$$

where we have assumed that $n_\mathrm{T} \gg n_\mathrm{nuc}$ so that $n_\mathrm{T} - n_\mathrm{nuc} \approx n_\mathrm{T}$. The quantity f_assem is the association rate constant for subunit addition, averaged over the growth phase since it may vary with aggregate size. The factor in the numerator indicates that the growth timescale generically increases with optimal aggregate size (i.e., $\alpha > 0$) since $\mathcal{O}(n_\mathrm{T})$ independent subunit additions must occur. The value of the exponent α will depend on factors such as the dimensionality, the aggregate geometry, and the relative stability of the intermediates. For example, for growth of a globular aggregate in 3D (i.e., a crystal or the SALR system of Sec. III.B.2) we expect $\alpha \approx 1/3$ if growth is proportional to the diffusion limited rate. For the capsid example that we consider in this section, we expect $1/2 \leq \alpha \leq 2$. In particular, if assembly is strongly biased over disassembly during growth and the assembly rate is proportional to the perimeter of the free boundary, we obtain $\alpha = 1/2$ (Hagan and Elrad, 2010). For moderately biased assembly, growth tends to occur along a single point on the perimeter, giving $\alpha = 1$, while for weakly biased assembly (near the reversible limit) the growth timescale approaches that of a random walk, giving $\alpha = 2$.

3. Beyond nucleation and growth

We note that not all assembly processes can be adequately described by the nucleation and growth mechanism. In some systems there are additional timescales that may become rate limiting. These include subunit conformational changes and cooperative global rearrangements required to achieve the optimal self-limited structure. For example, strains must propagate across scales of the order of the size of the structure in the frustrated open-boundary assemblies described in

---

[24] However, even for a system with no minimum in its aggregation free energy, one can define a growth timescale to reach a given finite size, which can be analyzed as described here.

Sec. III.B. Similar behaviors may occur in self-closing aggregates; for example, recent evidence suggests that assembly of empty HBV capsids proceeds by nucleation, growth into large but defective or disordered intermediates, followed by a "completion phase" in which the intermediate rearranges into the icosahedral capsid structure (Chevreuil *et al.*, 2020). Additional multistep mechanisms are possible if we move beyond the scope of this review (SLA from a single species). For example, computation and experiments have shown that assembly around a substrate or template can proceed by an alternative *en masse* pathway (Hagan, 2008; Elrad and Hagan, 2010; Tsvetkova *et al.*, 2012; Garmann *et al.*, 2014, 2016; Perlmutter, Perkett, and Hagan, 2014; Panahandeh *et al.*, 2020). In this process, subunits rapidly adsorb on a substrate in a disordered manner, then cooperatively rearrange to form an ordered aggregate. While that process involves condensation and assembly of individual substrates, it is also possible for the assembling components to undergo bulk phase separation into a metastable liquid phase prior to assembly. For example, in many virus families the host cell undergoes liquid-liquid phase separation to form a domain that is concentrated in viral proteins and nucleic acids, within which the nucleocapsid (capsid assembled around the viral RNA) assembles (Nikolic *et al.*, 2017; Schoelz and Leisner, 2017; Brocca *et al.*, 2020; Carlson *et al.*, 2020; Fernández de Castro, Tenorio, and Risco, 2020; Guseva *et al.*, 2020; Kieser *et al.*, 2020; Savastano *et al.*, 2020). A related mechanism occurs for unlimited assembly: it has been shown that crystallization can proceed with use of a two-step mechanism in which subunits first condense into a metastable liquid phase and formation of an ordered crystal follows; see ten Wolde and Frenkel (1997), Nicolis and Nicolis (2003), Gliko *et al.* (2005), Veesler *et al.* (2006), Fortini, Sanz, and Dijkstra (2008), Basios *et al.* (2009), Sear (2009), Whitelam (2010), and Schubert *et al.* (2017).

B. Interplay between thermodynamic stability, assembly rates, and kinetic traps

Achieving productive assembly requires thermodynamic stability of the target structure, which implies that subunit interactions must be strong enough to overcome the translational and rotational entropy losses incurred by the subunits forming an aggregate. Achieving productive assembly in finite time places even more restrictive conditions on interactions, based on the kinetics of assembly. Interactions must be strong enough to ensure that the previously discussed nucleation timescale falls within relevant timescales. However, overly strong subunit interactions lead to *kinetic traps*, or metastable states that evolve toward equilibrium slowly. These kinetic traps can be broadly classified into two categories. We discuss each category and its effects in turn and follow up with a survey of open questions for optimizing assembly kinetics.

1. Overnucleation (i.e., monomer starvation)

First, the *monomer starvation* trap arises when nucleation timescales are short in comparison to growth timescales, or the time required for a nucleated aggregate to grow to its equilibrium size. In this situation, so many nuclei form that the system becomes depleted of monomers before most nuclei grow to completion. Subsequent evolution to equilibrium requires either redistribution of subunits from smaller to larger aggregates (Ostwald ripening), which incurs significant free energy barriers, or coalescence of large intermediates, which is rare and frequently leads to misassembled structures.

In the context of SLA, this condition becomes more stringent as the target assembly size increases since about n_{T} monomers will eventually be depleted during the growth of each nucleus, and the growth timescale typically increases with target size since the critical nucleus size depends at most weakly on n_{T}. The parameter regimes that give rise to the monomer starvation trap can be understood from the different dependence of nucleation and growth timescales on subunit-subunit interaction strengths and subunit concentrations (Zlotnick *et al.*, 1999; Endres and Zlotnick, 2002; Hagan and Elrad, 2010). In particular, from Eq. (67) the initial nucleation timescale for the capsid model is $\tau_{\mathrm{nuc}}(\Phi) \sim \exp\{-\beta\Omega[n_{\mathrm{nuc}}(\Phi)]\}$. However, as previously noted the nucleation timescale increases as the reaction proceeds due to monomer depletion. By evolving aggregates according to the kinetics described by Eq. (C1), we can integrate the cumulative depletion of monomers as a function of time, from which one can obtain the following median assembly time $\tau_{1/2}$, which is defined as the time required for half of the subunits to be assembled $\phi_1(\tau_{1/2}) = \Phi/2$ (Hagan and Elrad, 2010) (see Appendix C for details):

$$\tau_{1/2} \sim \frac{\tau_{\mathrm{nuc}}(\Phi)}{\Phi n_{\mathrm{T}}}. \tag{72}$$

The boundary between productive assembly (which we define as having initial nucleation times shorter than, say, one day) and monomer starvation can be estimated by the locus in parameter space at which the median assembly time and growth time are equal. Figure 19 shows this and boundaries between other kinetic regimes as a function of supersaturation and n_{T} for the previously described capsid model. For any optimal size, we see that, as the total subunit concentration increases, the system transitions from a monomer-dominated equilibrium phase to a state in which the monomer rich phase is metastable only with respect to aggregates, but assembly does not occur on relevant timescales due to a large nucleation barrier, to a window of productive assembly, and finally to the monomer starvation regime. Notice that the concentration at which assembly becomes kinetically accessible can significantly exceed the CAC, showing that accurately inferring the CAC from experimental measurements can be challenging. Moreover, the region of productive assembly between the nucleation threshold and the monomer starvation trap narrows with increasing target size due to the different dependencies of the nucleation and elongation timescales on n_{T}. In particular, the median assembly time increases with n_{T} at small sizes due to the previously discussed rim curvature but decreases as $1/n_{\mathrm{T}}$ at large sizes because the nucleation time saturates while each nucleus consumes n_{T} subunits; see Eq. (72). Further, the growth time increases with n_{T} [Eq. (71) with $\alpha = 1$].[25]

---

[25] Note that a similar analysis can be performed for crystallization in a finite-sized system to tune nucleation and growth rates so that a single nucleus forms on accessible timescales but grows to system size before additional nuclei arise.

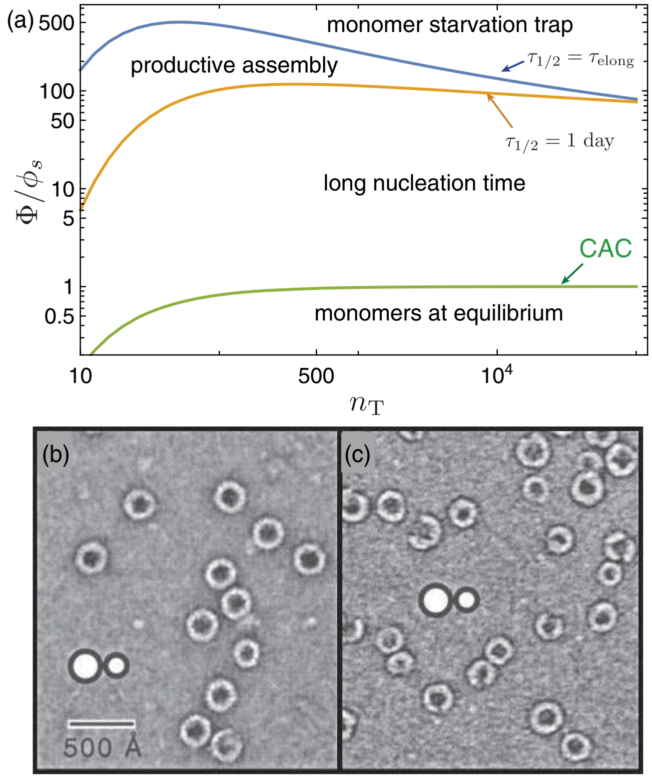

FIG. 19. (a) Assembly "phase diagram" for the capsid model ($B \to \infty$ model of the fluid capsule model). The boundaries between the different kinetics regimes discussed in the text are shown as a function of capsid size and supersaturation $\Phi/\Phi_{\mathrm{s}} = e^{-\beta\Delta\mu}$. The calculation was performed using Eqs. (68), (5), and (71) with $\alpha = 1$, as well as Eq. (C2). We set $\epsilon_{\min} = -15k_BT$ and $\lambda = \epsilon_{\mathrm{T}}/2a_0^{1/2}$ as in Fig. 18. (b),(c) TEM images of *in vitro* assembly of empty capsids from cowpea chlorotic mottle virus capsid proteins. (b) corresponds to productive assembly, while (c) corresponds to assembly of long-lived partial shells (the monomer starvation trap) that occurs under stronger subunit-subunit interactions. From Zlotnick *et al.*, 2000.

2. Malformed assemblies

In the second class of kinetic traps, misassembled structures arise when incorrectly bound subunits do not have time to anneal before becoming trapped within the aggregate by binding of additional subunits. That is, subunit interactions must be reversible on the timescale of subunit association, even in the growth phase of aggregate assembly. This condition depends on both the cohesive interaction strength and initial monomer concentration since the annealing time increases exponentially with interaction strength, while the time interval between subunit association events decreases inversely with monomer concentration. Moreover, in the case of open-boundary assembly (and possibly some instances of curvature-controlled assembly), subunit association must occur sufficiently slowly and close to reversibility that strains due to frustration have time to propagate across the structure. Otherwise, defects or cracks that allow the system to escape frustration may occur. The conditions leading to such escape scenarios remain an open question in open-boundary assembly.

A similar trade-off between thermodynamic and kinetic considerations applies to the orientational specificity of subunit interactions. While some degree of orientational specificity is required to stabilize the target structure over other competing morphologies and to avoid misassembly, overly high specificity (such as extremely precise lock-and-key interactions) leads to small kinetic cross sections for subunit association and thus low assembly rates (Whitelam *et al.*, 2009).

3. Nonequilibrium protocols for optimal kinetics

Finally, we note that it may be possible to achieve faster assembly and/or higher assembly yields of SLA over a larger region of parameter space by using nonequilibrium protocols in which assembly driving forces are varied over time, for example, by varying the temperature, solution conditions, or concentration. An intuitive approach is to use strong subunit interactions for sufficient duration to rapidly form a desired number of nuclei, then to reduce subunit interactions to a level where postnuclei intermediates can undergo growth but further nucleation is suppressed by a large nucleation barrier. Several approaches have been developed to use feed forward or feedback control to optimize such time varying protocols (Klotsa and Jack, 2013; Tang, Rupp *et al.*, 2016; Tang *et al.*, 2017; Pineros *et al.*, 2018; Green *et al.*, 2019; Grover, Griffin, and Tang, 2019). A simpler but highly effective strategy is to seed nucleation of particular structures (Mohammed and Schulman, 2013). This is related to the more general context of harnessing nonequilibrium assembly pathways to achieve size-controlled aggregates *out of equilibrium*, which we return to in detail in Sec. V.

V. FINITE SIZES BY OTHER MEANS

In Secs. II.B and III, we overviewed the statistical thermodynamics of known physical mechanisms for identical subunits to achieve equilibrium states with well-defined and finite dimensions. Here we survey two broad classes of mechanisms for achieving size-controlled assemblies that fall outside of this basic paradigm: nonequilibrium size control and programmable assembly multiple species.

A. Nonequilibrium mechanisms of size-controlled assembly

While we showed in Sec. IV that kinetic effects introduce limitations to the practical ability to achieve equilibrium SLA, in this section we consider how nonequilibrium effects can be exploited to achieve size control of finite assemblies even when the equilibrium states are non-self-limiting.

Such nonequilibrium mechanisms leading to size-controlled assembly distributions can be further classified into two categories. (i) In kinetically controlled assembly reactions, kinetic effects drive a system to a well-defined metastable state that either is sufficiently long-lived for practical applications or can subsequently be stabilized by additional reactions. Examples of such kinetically controlled reactions include polymer-particle synthesis processes (de Pablo *et al.*, 2019), nonequilibrium formation of finite-size droplets in microemulsions (Woltornist *et al.*, 2015, 2017), flow-driven aggregate breakup (Conchúir and Zaccone, 2013) and kinetically arrested coarsening (Siggia, 1979). (ii) The second class, typically referred to as nonequilibrium assembly mechanisms, requires continual energy input into the system to stabilize the self-limited size distribution. These mechanisms use energy consumption to modify the aggregate size dependence of subunit association and/or dissociation rates. In both equilibrium and nonequilibrium mechanisms, a stable assemblage requires association and dissociation rates to be equal at a finite aggregate size. The resulting fixed point must also be stable, requiring dissociation rates to exceed association rates at larger aggregate sizes and association rates to be larger for smaller aggregates. These conditions are guaranteed in an equilibrium system by the criteria for self-limited assembly sizes discussed in Sec. II.B.2. However, additional nonequilibrium mechanisms to modify association or dissociation rates can also lead to stable finite sizes. For example, a number of biological mechanisms have been proposed in which assembly sizes are regulated by energy-consuming processes, such as active assembly and disassembly by molecular machines, or coupling between protein conformational states and phosphotransfer reactions (changes in a protein's phosphorylation state or hydrolysis of a bound, nucleotide triphosphate, etc.). Similarly, assembly and disassembly rates can be modulated by externally controlled gradients in monomer concentrations or nucleation factors, as occurs during embryogenesis (Briscoe and Small, 2015).

In a well-studied example, the *lengths* of microtubules or actin filaments are regulated by a suite of accessory proteins that modulate assembly and disassembly rates at filament ends, as well as molecular motors that actively remove subunits from filament ends or sever filaments in their interiors (Desai and Mitchison, 1997; Mohapatra *et al.*, 2016; Pollard, 2016). Moreover, the subunits themselves undergo conformational changes powered by hydrolysis of nucleotide triphosphates [adenosine triphosphate (known as ATP) or guanosine triphosphate (GTP)]. The hydrolysis event shifts subunits into geometries that are incompatible with the global filament structure, thus inhibiting further assembly and/or weakening the existing structure. These nonequilibrium processes allow not only stable finite-length distributions but also structural dynamics, such as actin treadmilling and microtubule dynamical instability, that allow the cellular cytoskeleton to rapidly respond and reconfigure to environmental cues. Using models ranging from idealized one-dimensional filaments to geometrically realistic particle-based dynamical simulations (Mohapatra *et al.*, 2016; Bollinger and Stevens, 2018; Fai *et al.*, 2019; Tong and Voth, 2020; Hemmat and Odde, 2021), researchers have identified multiple mechanisms by which active filament assembly and disassembly processes and energy-driven subunit conformational changes can lead to 1D filaments that exhibit dynamical instabilities and/or have well-defined stable sizes. In contrast, recall from Sec. II.A.2 that equilibrium 1D filaments generically exhibit exponential length distributions.

Other biological structures thought to be subject to nonequilibrium size regulation include coat protein (known as COP) bound vesicles in the eukaryotic secretory system (Foret and Sens, 2008), neuronal synapses (Lisman and Raghavachari, 2006, 2015; Burlakov *et al.*, 2012; Broadhead *et al.*, 2016; Liu, Hagan, and Lisman, 2017; Miermans *et al.*, 2017; Shomar *et al.*, 2017), transcriptional regulatory complexes (enhancers) (Hnisz *et al.*, 2017; Cho *et al.*, 2018; Chong *et al.*, 2018; Sabari *et al.*, 2018), and other phase-separated liquid domains (Zwicker *et al.*, 2014; Zwicker, Hyman, and Julicher, 2015; Weber *et al.*, 2019). The key characteristic of all of these systems is that the finite size of the assembled structure depends on continual energy consumption; e.g., through the replacement of "inactive" subunits with "active" ones using the disassembly of GDP-bound tubulin subunits and reassembly of GTP-bound tubulin at a microtubule end, or the continual dephosphorylation and rephosphorylation of subunits by phosphatases and kinases within a neuronal synapse (Lisman and Raghavachari, 2006, 2015).

More broadly, it has been known since Turing's seminal paper (Turing, 1952) that combining imbalances in diffusion rates with interconverting molecular species (such as through chemical reactions) can lead to compositional inhomogeneities with well-defined steady-state sizes (Haselwandter *et al.*, 2011, 2015; Halatek and Frey, 2018). More recently advances in stochastic thermodynamics (Seifert, 2008, 2012) have demonstrated that active processes can play important roles in regulating the structures and functions of assembly and self-organization (Nguyen and Vaikuntanathan, 2016; Marsland and England, 2018). Moreover, theoretical and experimental studies suggest that spatiotemporal patterns with well-defined domain sizes can occur in some active matter systems whose constituent components consume energy at the particle scale to drive motion (Marchetti *et al.*, 2013; Bechinger *et al.*, 2016; Needleman and Dogic, 2017; Doostmohammadi *et al.*, 2018; Bär *et al.*, 2020; Shaebani *et al.*, 2020).

B. Addressable assembly of programmable subunit mixtures

In Sec. III, we reviewed two broad categories of self-limiting assemblies, both of which achieve equilibrium finite-size assemblies from a single species of subunit. Here we describe an emerging class of self-assembling systems that also realizes equilibrium finite-size assembly but which falls (at least partly) outside of these two categories.

We refer to this class of systems as *addressable assemblies* (AAs) following the terminology introduced by Jacobs, Reinhardt, and Frenkel (2015, 2016). AAs are formed by mixtures of multiple assembling subunit species (species A, B, C, D, etc.), each with specific interactions that selectively bind to a subset of all subunit species (e.g., A binds selectively only to D, while B binds to itself as well as C). The core design principle of AAs is this: one can "program" the matrix of species interactions to match the 3D adjacency matrix of a desired structure, perhaps uniquely, such that this target structure becomes the equilibrium assembly state in mixtures of controlled subunit stoichiometry (Hormoz and Brenner, 2011). That is, each particle has an address (or set of addresses) where it sits in the 3D target assembly.

Examples of AAs include colloids and nanoparticles functionalized by single-stranded DNA tethers that mediate interactions via complementary base pairing (Jones, Seeman, and Mirkin, 2015), as well as DNA bricks assembled by mixtures of oligomeric DNA strands whose sequences are designed to interleave ends into 3D patterns via complementary base pairing (Ke *et al.*, 2012). Each of these systems has been considered and studied for its potential to selectively design and assemble target superstructrures that terminate at specifically predetermined dimensions (Zeravcic, Manoharan, and Brenner, 2014). In the sense that the self-limiting target structure can be designed to be stress free with each cohesive bond ideally satisfied, finite AAs can be thought of as multispecies analogs to self-closing structures, albeit with more complex bond networks. As an example, rectangular beams with precisely defined and finite cross-sectional dimensions were assembled using the programmable DNA bricks (Ke *et al.*, 2014). The different values of the target finite dimension each required mixtures composed of different numbers of distinct oligomeric species. For example, finite-width beams of 6×6, 8×8, and 10×10 double-stranded DNA helices across were assembled, respectively, from mixtures containing 60, 112, and 180 distinct oligomers. Generically, equilibrium termination at specific size via AA requires a number of distinct subunit species $n_{\rm S}$ that grows with the finite target size W_* (Ong *et al.*, 2017), presumably with some power law $n_{\rm S} \sim W_*^{\beta}$ (e.g., $\beta = 2$ if distinct subunits are required in every 2D cross section of a beam).

The unbounded growth of the number of subunit species with target size for such an implementation of AA would seem to limit its practical applicability due to the cost (in terms of design, synthesis, and processing) and limited scalability of the programmable mixture. This raises basic questions about AA: Are there optimal strategies that minimize the complexity (such as the number of distinct species) of a subunit mixture needed to achieve self-limiting AAs of a given topology, and how do these scale in the limit of large target size? The self-assembly of viral shells, or capsids, may provide clues for how to approach these questions.

Capsids are "crystalline" shells composed of protein subunits (capsomers) that self-assemble to enclose the viral genome. Since the enclosed genome has to code for the capsomers themselves, it has been understood since Crick and Watson's seminal paper (Crick and Watson, 1956) that such assemblies should be economical. That is, a viral capsid is under selective pressure to enclose the largest possible volume using the minimal number of distinct capsomer types. Caspar and Klug (CK) proposed that viruses achieve this optimization by exploiting symmetry principles (Caspar and Klug, 1962). In their well-known construction, quasispherical capsid structures are mapped to high-symmetry triangulations of the sphere, in which each triangle is constituted of three capsomeric subunits (Prasad and Schmid, 2012). Based on this reasoning, CK conjectured that optimal capsids correspond to subtriangulations of the icosahedron. Each CK structure can be classified by the number $T = h^2 + k^2 + hk$ of subtriangles in each of the 20 triangular faces of the icosahedron, where h and k are positive integers. Since regular triangulations of the sphere are not possible beyond those corresponding to Platonic solids, CK noted that higher triangulation numbers require capsomers to accommodate different local environments corresponding to variations in neighbor spacing, orientation, and the number of neighbors. In CK capsids, the triangulation number T is the number of symmetry-inequivalent capsomer positions.

In this way, we might view CK capsids as a highly symmetric and economical limit of AA. While the complete shell could be assembled from $60T$ distinct and specifically interacting capsomeric subunits, the large number of symmetry elements of the icosahedral net implies that there are many redundancies in such a design, and in fact the same unique target structure could be realized from specific interactions of only T distinct units. This approach has been adopted by the protein design community, enabling researchers to engineer proteins that assemble into icosahedral shells of various sizes (King *et al.*, 2014; Lai *et al.*, 2014; Bale *et al.*, 2016; Butterfield *et al.*, 2017; Mosayebi *et al.*, 2017). Recently the CK design principles have been repurposed for the *de novo* design of triangular DNA origami particles with precisely defined geometry and edge interactions that selectively assemble into T icosahedral shells (Sigl *et al.*, 2021). They show that such capsids can be uniquely programmed and assembled from even a fewer number of distinct subunit species, $\lceil T/3 \rceil$ because each triangle can have three inequivalent edges.

In its simplest implementation, a CK capsid can be assembled by synthesizing a distinct subunit species for each of the T symmetry-equivalent positions. While some small viruses follow this approach, it becomes increasingly impractical as the target capsid size grows since the number of distinct species scales linearly with the capsid area. In practice, there are two mechanisms to reduce the number of distinct subunits that need to be synthesized by a virus (or by any other manufacturer). In many viruses, the capsid protein interconverts between different "quasiequivalent" conformations with slightly different interaction geometries that accommodate the different local symmetry environments in the capsid, thus enabling assembly of capsids with $T > 1$ from a single-subunit species (Caspar and Klug, 1962; Johnson and Speir, 1997). More broadly, it has been shown that such high-symmetry constructions correspond to free energy minima of assembled structures with spherical topologies with relatively generic types of short-range cohesive subunit-subunit interactions, although the free energy minimum symmetry depends on the size of the assembled structure (Zandi *et al.*, 2004; Chen, Zhang, and Glotzer, 2007). Correspondingly, CK-like capsids emerge naturally from the assembly of elastic structures from subunits resembling the tapered subunits discussed in Sec. III.A.2 (Chen, Zhang, and Glotzer, 2007; Fejer, Chakrabarti, and Wales, 2010; Lázaro, Dragnea, and Hagan, 2018; Lázaro, Mukhopadhyay, and Hagan, 2018; Reguera, Hernández-Rojas, and Gomez Llorente, 2019), and many of the principles discussed in that section can be extended to systems that form CK capsids. However, the assembly dynamics of such structures remains an open question: how does the right subunit conformation end up in the appropriate location within an assembling capsid (Berger *et al.*, 1994; Elrad and Hagan, 2008; Morton *et al.*, 2010; Stockley, Ranson, and Twarock, 2013; Perlmutter and Hagan, 2015; Perkett, Mirijanian, and Hagan, 2016; Li *et al.*, 2018; Panahandeh, Li, and Zandi, 2018; Twarock *et al.*, 2018; Panahandeh *et al.*, 2020; Zandi *et al.*, 2020)? More recently the CK construction has been extended to account for other viruses in which capsomers accommodate more extreme differences in local environments and other symmetry classes (Twarock, 2004; Luque and Reguera, 2010;

Twarock and Luque, 2019). In a second strategy, finite shells are assembled from a single (or a few) subunit types that do not adopt explicitly distinct conformations but are sufficiently deformable to accommodate different local symmetry environments by the formation of inhomogeneous strains within the capsid structure. It is easy to imagine that these two strategies (different conformations with specific interactions or subunits with deformable interaction geometries) could be combined to extend the size and complexity of a shell that can be assembled by the CK mechanism.

Returning to the context of AA, CK constructions and their extensions can be viewed as limiting cases of AA in which one uses symmetry to putatively minimize the number of distinct subunits types need to enclose a given volume. The CK framework suggests that there is a trade-off between the complexity of the mixture of distinct subunits and the complexity, or asymmetry, of the target assembly (i.e., fewer symmetry elements in the target structure implies more subunit species). This idea can be extended more broadly to designing self-limiting AAs that target other, nonspherical topologies. In this context, a more precise and potentially useful notion of economy may be to consider the ratio of the target finite size to the number of subunit types W_*/n_{S}, and ask what the analogs to CK designs for arbitrary topologies are that maximize this ratio in the limit of increasing W_*.

VI. CONCLUDING REMARKS: SELF-LIMITING ASSEMBLY BY DISCOVERY AND DESIGN

In this review we attempted to provide a unified theoretical perspective on assembly processes that, while occurring in chemically and physically diverse systems, share the common thread of autonomously terminating at a well-defined equilibrium finite size. Starting from the framework of ideal aggregation theory, we showed that the necessary conditions for such self-limited assembly are linked to the existence of a minimum in the size-dependent aggregation energetics $\epsilon(n)$. In contrast, systems that do not meet these conditions generically assemble bulk structures such as crystals. We saw that properties of self-limited assembly reactions can be identified from the functional form of $\epsilon(n)$, including the onset of aggregation with increasing subunit concentration and the size of fluctuations around the optimal aggregate size. Systems with multiple local minima in $\epsilon(n)$ can exhibit polymorphic self-limited assembly, with concentration-dependent transitions (secondary CACs) between different aggregation states. We saw that the small, but nonzero, translational entropy of the aggregates plays a key role in driving secondary CACs.

The existence of local minima at nontrivial sizes $n_{\mathrm{T}} \gg 1$ implies that the energetics of assembly, as mediated by the shape and interactions between the subunits themselves, requires the ability to sense the aggregate size on scales comparable to the optimal size. Surveying known examples of equilibrium SLA from identical subunits, we argued that there are two broad classes of physical mechanisms that achieve this. Either the subunit interaction geometries encode a target assembly geometry that "returns to itself" after a characteristic number of subunits or, instead, there is a source of intra-assembly stress gradients that can propagate up to the finite-size scale of the assembly. An important feature of the latter mechanism is that it enables finite-sized aggregates that have open boundaries. We presented generic descriptions for understanding each of these mechanisms, including how to assess the limits to the finite size that can be achieved. Toward that end, we considered mechanisms by which the system can escape finite size, resulting in bulk assembly products. For example, a system can curtail the accumulation of stress gradients with assembly size by expelling strain to its periphery or by locally relaxing strain through defects. We presented distinct physical mechanisms that can give rise to accumulating stress gradients: the interplay between short-range attractions and long-range repulsions or an incompatibility between the preferred local subunit packing and the large-scale assembly geometry. We also presented examples of experimental systems that may correspond to each of these SLA mechanisms.

In part, our purpose in spotlighting the relatively rarefied conditions required for equilibrium self-limitation is to reframe broadly open challenges in understanding and engineering SLA. One such challenge could be described as the experimental inference of SLA. That is, for a given set of experimental observations of assembly, is it possible to determine whether finite aggregates are the result of equilibrium self-limitation? This is a particularly vexing issue for experimental observations of both synthetic and biological assemblies that appear to be finite and well defined. While it is often desirable to link the observations to specific microscopic models that recapitulate aspects of finite assembly *a posteriori*, such models often require assumptions about the interactions and energetics of complex subunits that are poorly understood. Thus, it is challenging to rule out alternative mechanisms of kinetic trapping of assembly in such systems. As one example, in living tissues collagen forms fibrillar assemblies that appear to have well-defined diameters reaching up to microns, well beyond the nanometer-scale width of a single procollagen molecule (Ottani *et al.*, 2002). Moreover, the mean diameter varies considerably between tissue types. Fibers in tendons have mean diameters in excess of $\gtrsim 1\ \mu$m, while those found in corneal tissue have a tighter distribution of around $\lesssim 50$ nm (Wess, 2008). Such observations combined with the functional needs of these different tissues (Meek, 2009), high stiffness versus optical transparency, respectively, suggest a need to regulate the assembly of the same subunits to form architectures of *tunable* finite size. Indeed, physical models have proposed mechanisms of geometric frustration deriving from the chiral organization within fibers as a means of imposing a self-limited diameter (Turner *et al.*, 2003; Grason and Bruinsma, 2007; Yang, Meyer, and Hagan, 2010; Brown, Kreplak, and Rutenberg, 2014). While these explanations are plausible and broadly consistent with observations of finite-diameter collagen fibers and other fibrous biofilament structures, these physical models require knowledge of parameters describing chiral intermolecular forces and intermolecular mechanics, which are difficult to predict under conditions relevant to assembly (Grason, 2020). Without direct knowledge of these intermolecular and intramolecular parameters, not to mention the nonequilibrium conditions of assembly, at best such models can plausibly explain the finite size of fiber diameters.

Developing generalizable experimental methodologies that strictly prove (or disprove) mechanisms of equilibrium self-limitation, particularly from complex biomolecular subunits for which the intermolecular aggregation energetics is poorly understood, remains a more distant and unmet goal.

A related challenge is to use advances in synthetic techniques to design and engineer self-limiting assemblies that target *a priori* finite dimensions. Great advances have been made in designing shape-controlled particles (Glotzer and Solomon, 2007; Sacanna and Pine, 2011) whose symmetries and interactions direct assembly to targeted structures. However, assembly targets have thus far been largely restricted to various bulk structures (albeit with complex unit cells), or 1D or 2D aggregates of uncontrolled ultimate size. Alternately, the field of supramolecular chemistry has leveraged chemical synthesis of a variety of architecturally and compositionally defined macromolecules to direct their assembly. The chemical control over these precision amphiphiles has significantly increased one's ability to rationally design and form periodic mesophases, thermotropic supramolecular crystals, or liquid crystals (Su *et al.*, 2020). However, while the symmetries of these phases have become increasingly complex, they remain bulk structures. Synthetic advances in amphiphile assembly have largely focused on imbuing micellar assemblies with functional properties, such as the controlled uptake and release of drugs (Geng *et al.*, 2007; Oltra, Nair, and Discher, 2014). While micellar assemblies are finite size in terms of diameter, as in the case of traditional surfactants, the finite size remains limited to the size of the molecules that span the aggregate core. Thus, notwithstanding tremendous advances in synthesizing shape- and interaction-controlled subunits, controlling the finite size of target superstructures, particularly on size scales much larger than the subunits themselves, remains a relatively unexplored aspect of engineered assemblies. While recent advances in methods such as DNA nanotechnology seem to pave the way for geometric control of subunits needed to realize bioinspired capsules and tubules (Rothemund *et al.*, 2004; Tian *et al.*, 2014; Benson *et al.*, 2015; Sigl *et al.*, 2021), it remains to be explored what the experimentally realizable upper limits to finite sizes are, and what mechanisms of self-limitation are needed to reach this limit.

A further challenge is to design supramolecular structures that do not have a single finite size but rather can exist in multiple different sizes, all of which are stable. Such classes of structures enable essential functions in biology. For example, recent evidence suggests that a neuronal synapse changes size during long-term memory storage but then must remain stable at that size over the lifetime of a memory (Lisman and Raghavachari, 2006; Tang, Chen *et al.*, 2016). Designing such variable-size stable structures is also becoming of interest to nanomaterials science since materials capable of learning or remembering multiple stable configurations (Murugan, Zou, and Brenner, 2015; Zhong, Schwab, and Murugan, 2017) could adapt their structures to store information, self-heal, or respond to environmental cues. Despite this interest and insights from biology, the principles underlying such variable-size stable structures remain far from clear. In this context, it would be of interest to extend the considerations in Sec. II.B.3 of secondary CACs to understand more broadly how the interplay between aggregate translational entropy and interaction energies can lead to controllable transitions between structures with different finite number sizes and/or dimensionalities. Similarly, can these principles be combined with the concepts of nonequilibrium assembly to design subunits that are preprogrammed to organize into nanoscale machines capable of autonomously manipulating matter or performing other functions currently found only in living organisms?

ACKNOWLEDGMENTS

The authors are grateful to P. Charbonneau, E. Efrati, B. Gelbart, N. Hackney, D. Hall, R. Jack, M. Lenz, B. Rogers, L. Tsidilkovski, B. Tyukodi, R. Zandi, and A. Zlotnick, and for valuable discussions and useful feedback on this manuscript. Support for this work was provided through the NSF through the the Brandeis Center for Bioinspired Soft Materials, a NSF MRSEC, Grants No. DMR-1420382 and No. DMR-2011846 (to M. F. H. and G. M. G.), and Grant No. DMR-2028885 (to G. M. G.), and the NIH through Grant No. R01GM108021 from the National Institute of General Medical Sciences (to M. F. H).

APPENDIX A: POLYMORPHIC AMPHIPHILE ASSEMBLY PHASE DIAGRAM

Here we summarize the calculation of the polymorphic assembly phase diagram for the amphiphile model in Fig. 12. The aggregate energy for each of the dimensionalities is given by Eq. (44). To determine the phase diagram, we calculate the following law of mass action for subunit populations:

$$\Phi = \phi_1 + \Phi_3 + \Phi_2 + \Phi_1, \tag{A1}$$

with Φ_{d_L} the mass fraction of subunits in spherical, cylindrical, or planar aggregates for $d_\mathrm{L} = 3, 2, 1$, respectively, and ϕ_1 the free monomer population

Adopting the continuum limit, the mass fraction of subunits in spheres is given by

$$\begin{aligned}\Phi_{3,\mathrm{cont}} &= \int_0^\infty dr 4\pi r^2 n_0 \phi_3(r),\\ \phi_3(r) &= n_\mathrm{sph}(r)\exp\{-\beta[\epsilon_\mathrm{sph}(r)-\mu]n_\mathrm{sph}(r)\},\end{aligned} \tag{A2}$$

with $\epsilon_\mathrm{sph}(r) = \epsilon(r,3)$ from Eq. (44), $n_\mathrm{sph}(r) = (4/3)\pi r^3 n_0$, and $n_0 = v_0^2/a_0^3$ a dimensionless number density.

Likewise, the mass fraction in spherocylinders is

$$\Phi_2 = \int_{L_\mathrm{B}}^\infty dL \int_0^\infty dr 2\pi r n_0 \phi_2(r,L)$$

$$\begin{aligned}\phi_2(r,L) = {}&[n_\mathrm{sph}(r)+n_\mathrm{cyl}(r,l)]\\ &\times\exp\{-\beta[\epsilon_\mathrm{SC}(r,L)-\mu][n_\mathrm{sph}(r)+n_\mathrm{cyl}(r,L)]\},\end{aligned} \tag{A3}$$

with L_B set by the minimal length of the stable spherocylinder branch (calculated later), $n_\mathrm{cyl}(r,L) = n_0\pi r^2 L$, and

$$\epsilon_{\mathrm{SC}}(r,L)=\frac{n_{\mathrm{sph}}(r)\epsilon_{\mathrm{sph}}(r)+n_{\mathrm{cyl}}(r,l)\epsilon_{\mathrm{cyl}}(r)}{n_{\mathrm{sph}}(r)+n_{\mathrm{cyl}}(r,L)},\qquad \text{(A4)}$$

with $\epsilon_{\mathrm{cyl}}(r)=\epsilon(r,2)$ from Eq. (44). Note that the first term in the numerator of Eq. (A3) corresponds to the end cap energy arising from the hemispherical cap at either end of the spherocylinder. We make the simplest assumption, that the radius of the hemispherical cap is equal to that of the cylindrical portion of the micelle, so that solvophobic tails remain shielded from solvent contact at the cylinder–end cap connection. More realistic models consider lower-energy shapes that smoothly connect bulbous ends to cylindrical cores (May and Ben-Shaul, 2001).

Finally, Φ_1 is the mass fraction in layers and is calculated later.

To proceed, recall from Sec. II.B.2 that, for the case of a minimum in the aggregate size around an optimal size n_{T}, fluctuations vanish in the limit of large n_{T} or $\epsilon''|_{n_{\mathrm{T}}}$. Specifically, consider a generic aggregate energy function $\epsilon(n)$ with a minimum ϵ_* at the optimal size n_{T}. The mass fraction of subunits in aggregates is then given by

$$\Phi_{\mathrm{T}}=\int_0^\infty dn\, n\, \exp\{\beta[\mu-\epsilon(n)]\}.\qquad \text{(A5)}$$

Performing a saddle point as in Sec. II.B.2 then results in

$$R_{\mathrm{fluc}}=\frac{\Phi_{\mathrm{T}}}{\Phi_{\mathrm{T}}(n_{\mathrm{T}})}=\sqrt{\frac{2\pi}{\beta\epsilon''|_{n_{\mathrm{T}}}n_{\mathrm{T}}}},\qquad \text{(A6)}$$

with $\Phi_{\mathrm{T}}(n_{\mathrm{T}})=n_{\mathrm{T}}e^{-\beta(\epsilon_*-\mu)n_{\mathrm{T}}}$ the mass of subunits in aggregates if fluctuations are neglected. Thus, we see that when $\beta\epsilon''|_{n_{\mathrm{T}}}n_{\mathrm{T}}\gtrsim 1$ the width of the distribution is smaller than a subunit, and fluctuations are negligible. For spherocylinders, we see that the contribution due to fluctuations in the radial direction r diminishes with length as $R_{\mathrm{fluc}}\sim L^{-1/2}$. Thus, even for narrow spherocylinders, fluctuations become negligible in the large-length limit.

With this in mind, we account for polydispersity in micelles as follows. We first calculate the optimal radius for spherical aggregates $\bar{r}_{\mathrm{sph}}$ by minimizing $\epsilon_{\mathrm{sph}}(r)$ for given values of k and P. We then numerically calculate the location of the barrier between the spherical and spherocylindrical branches of Eq. (A4), as $L_{\mathrm{B}}=\arg\max_L[\min_r\epsilon_{\mathrm{SC}}(r,L)]$, i.e., the length at which a spherocylinder of optimal radius has a maximum energy per particle. The number of particles at the barrier is then given by $n_{\mathrm{B}}=n_{\mathrm{sph}}(r_{\mathrm{B}})+n_{\mathrm{cyl}}(r_{\mathrm{B}})$, with $r_{\mathrm{B}}=\arg\max_r\epsilon_{\mathrm{SC}}(r,L_{\mathrm{B}})$ the optimal radius at the barrier.

To calculate the mass of spherical micelles, we numerically integrate the expression in Eq. (A2). To maintain the assumptions of Sec. III.A.2, we perform the integral over the range $n\in[\bar{n}_{\mathrm{sph}},n_{\mathrm{B}}]$, with $\bar{n}_{\mathrm{sph}}=n_{\mathrm{sph}}(\bar{r}_{\mathrm{sph}})$, although the result is largely insensitive to increasing these integration bounds. We then include fluctuations only when they exceed the size of a single subunit by setting

$$\Phi_3=\max\left[\Phi_{3,\mathrm{cont}},\bar{n}_{\mathrm{sph}}e^{-\beta[\epsilon_{\mathrm{sph}}(\bar{n}_{\mathrm{sph}})-\mu]\bar{n}_{\mathrm{sph}}}\right],\qquad \text{(A7)}$$

where the second argument is simply the concentration of micelles at the optimal size.

For spherocylinders we make the simplifying assumption that radial fluctuations can be neglected at all lengths and take the optimal radius $\bar{r}_{\mathrm{SC}}(L)=\arg\max_r\epsilon_{\mathrm{SC}}(r,L)$ as a function of spherocylinder length L. In practice, we found that the numerics are more tractable if the integral is performed over particle number n rather than spherocylinder length L, so we calculate the optimal spherocylinder length $L_{\mathrm{SC}}(n)=\arg\max_L\epsilon_{\mathrm{SC}}[\bar{r}_{\mathrm{SC}}(n,L),L]$, with $\bar{r}_{\mathrm{SC}}(\hat{n},L)$ determined from the volume of the spherocylinder $n_{\mathrm{sph}}(\bar{r}_{\mathrm{SC}})+n_{\mathrm{cyl}}(\bar{r}_{\mathrm{SC}},L)=\hat{n}$. To make the integral numerically tractable, we perform the integral to a predefined large size $n_{\max}$, beyond which we assume that the effect of changing radius from the hemispherical caps is negligible, so that the optimal radius is given by the minimum of the cylinder energy $\bar{r}_{\mathrm{SC}}\approx\bar{r}_{\mathrm{cyl}}=\arg\max_r\epsilon_{\mathrm{cyl}}(r)$, and the energetics becomes simply the 1D energetics of the form in Sec. II.A.2:

$$\Phi_2\approx\int_{n_{\mathrm{B}}}^{n_{\max}}dn\,\Phi_2\{\bar{r}_{\mathrm{SC}}[n,L_{\mathrm{SC}}(n)],L_{\mathrm{SC}}(n)\}+\Phi_\infty(n_{\max}),\qquad \text{(A8)}$$

with the contribution from the spherocylinders with sizes larger than $n_{\max}$ given by

$$\Phi_\infty(n_{\max})=e^{-\beta\epsilon_{\mathrm{cap}}}e^{\beta(\mu-\bar{\epsilon}_{\mathrm{cyl}})n_{\max}}\frac{1+\beta(\bar{\epsilon}_{\mathrm{cyl}}-\mu)n_{\max}}{\beta^2(\bar{\epsilon}_{\mathrm{cyl}}-\mu)^2},\qquad \text{(A9)}$$

with $\bar{\epsilon}_{\mathrm{cyl}}=\epsilon_{\mathrm{cyl}}(\bar{r}_{\mathrm{cyl}})$ the energy per particle within the cylindrical region and $\epsilon_{\mathrm{cap}}=n_{\mathrm{sph}}(\bar{r}_{\mathrm{cyl}})[\epsilon_{\mathrm{sph}}(\bar{r}_{\mathrm{cyl}})-\bar{\epsilon}_{\mathrm{cyl}}]$ the total extra energy that arises due to the unfavorable hemispherical caps.

Finally, the fraction of subunits in layers Φ_1 is calculated by noting that the free subunit chemical potential can never exceed the chemical potential of a subunit in a sheet $\phi_1\leq e^{\beta\bar{\epsilon}_{\mathrm{layer}}}$, with $\bar{\epsilon}_{\mathrm{layer}}=\min_r\epsilon_{\mathrm{layer}}(r)$, where $\epsilon_{\mathrm{layer}}(r)=\epsilon(r,1)$ from Eq. (44). Thus, the amount of subunits in layers and the corresponding free subunit concentration are given by mass conservation as

$$\begin{aligned}\Phi_1&=\max\left[\Phi-\Phi_3-\Phi_2-e^{\beta\bar{\epsilon}_{\mathrm{layer}}},0\right],\\ \phi_1&=\Phi-\Phi_3-\Phi_2-\Phi_1,\end{aligned}\qquad \text{(A10)}$$

effectively treating layers as an unlimited bulk phase, with negligible edge energy.

Phase boundaries in Fig. 12(a) are calculated from Eqs. (A2),(A3),(A4),(A5),(A6),(A7),(A8),(A9),(A10) by determining the total subunit concentration at which the fraction of subunits in an aggregate of a given dimensionality exceeds 50%. That is, the concentrations corresponding to transitions between monomers and layers, layers and spherocylinders, and spherocylinders and spheres are calculated as the lowest total concentration at which $\Phi(\Phi_1=0.5)$, $\Phi(\Phi_2=0.5)$, and $\Phi(\Phi_3=0.5)$, respectively. The concentrations in Fig. 12 are normalized by the CAC for cylinders ($\phi^*=e^{\beta\bar{\epsilon}_{\mathrm{cyl}}}$). Note that ϕ^* is the only result in this section that depends on the cohesive energy strength ϵ_0; the relative concentrations corresponding to the transitions depend only on the elastic and boundary energy terms.

The CAC ratio shown within the spheres and spherocylinders coexistence region in Fig. 12(b) is computed as $\Phi(\Phi_3 = 0.5)/\Phi(\Phi_2 = 0.5)$. The infinite-concentration transitions, shown as solid red and blue lines in Fig. 12(b), are calculated from $\bar{\epsilon}_{\text{layer}}(\bar{k}, P) = \bar{\epsilon}_{\text{cyl}}(\bar{k}, P)$ and $\bar{\epsilon}_{\text{cyl}}(\bar{k}, P) = \bar{\epsilon}_{\text{sph}}(\bar{k}, P)$, respectively.

Finally, the boundary of the spheres and spherocylinders coexistence region [the solid green line in Fig. 12(b)] is estimated as the minimum concentration at which either the height of the barrier between the sphere and spherocylinder branches goes to zero [i.e., corresponding to $\delta = 0$ in Eq. (34) of Sec. II.B.3] or the point at which the transition concentration from spheres to spherocylinders becomes equal to the threshold concentration for assembling cylinders in the absence of other aggregates [Eq. (35)].

APPENDIX B: CONTINUUM ELASTIC THEORY OF FRUSTRATED CHIRAL RIBBONS

Here we present details of the "narrow-ribbon" theory of frustrated chiral ribbons. As the model has been described elsewhere (Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014; Grossman, Sharon, and Diamant, 2016), our primary aim is to provide more details on the physical ingredients of the model, and further to describe how the elastic instability of wide helicoids quantitatively alters the picture of "frustration escape" presented in Sec. III.B.3.

Following the approach of Ghafouri and Bruinsma (2005), we consider a simplified theory that describes the shape of ribbons in terms of the surface curvature tensor C_{ij} along the midline of ribbons, written in terms of coordinate directions $\hat{x}_e$ and $\hat{x}_p$ that point, respectively, along and perpendicular to the wide direction of the ribbon; see Fig. 17(a). Specifically, this assumes that the in-plane curvatures vary slightly away from the midline of the ribbon, which is strictly valid when the ribbon widths are narrow with respect to their curvature radii. Since the ribbons effective flatten in shape as they grow wider, this narrow-ribbon approximation provides at least a qualitative picture of the ribbons' thermodynamics over the entire range of the widths.

The excess energy derives from two elastic contributions

$$E_{\text{elast}} = E_{\text{intrinsic}} + E_{\text{extrinsic}}, \tag{B1}$$

where the first term depends on the intrinsic geometry or metric distortions away from a planar 2D lattice, while the second term describes the elasticity of the extrinsic geometry of the ribbon, i.e., a generalized form of its bending energy expressed as quadratically in terms of curvature elements C_{ij}. The former term is captured by a 2D elastic energy

$$E_{\text{intrinsic}} = \int dA\, \sigma_{ij} u_{ij}, \tag{B2}$$

with u_{ij} and $\sigma_{ij} \approx Y u_{ij}$ as the respective in-plane stress and strain of the 2D crystal ribbon order, and Y as the 2D Young's modulus of the crystal (Seung and Nelson, 1988). Assuming that the crystalline packing favors uniform inter-subunit spacing, there is a geometrical and mechanical coupling between in-plane stress and out-of-plane deflections described by the so-called compatibility equation

$$\nabla_{\perp}^2 \sigma_{ii} = -Y K_{\text{G}}, \tag{B3}$$

where $\sigma_{ii} = \sigma_{pp} + \sigma_{ee}$ and $K_{\text{G}} \simeq C_{ee} C_{pp} - C_{ep}^2$ is the Gaussian curvature (neglecting variations of K_{G} across the ribbon width). Assuming approximately uniform negative curvature, it is straightforward to show that $E_{\text{intrinsic}}/WL = Y K_{\text{G}}^2 W^4/1440$ (Ghafouri and Bruinsma, 2005; Grason, 2016).

The extrinsic energy takes the form of the generalized bending energy

$$E_{\text{extrinsic}} = \frac{1}{2} \int dA [B_{pp} C_{pp}^2 + B_{ee} C_{ee}^2 + 2B_{ep}(C_{ep} - \Omega_0)^2], \tag{B4}$$

where B_{ij} are bending coefficients for different curvature elements. Here we consider the case of $B = B_{pp} = B_{ee} = B_{ep}$; relaxing this restriction does not alter the qualitative behavior. Symmetry considerations (i.e., lack of inversion symmetry) argue that chirality at the subunit scale generates a linear coupling to the off-diagonal curvature component, which we define as the ribbon twist

$$\Omega \equiv C_{ep}, \tag{B5}$$

and hence Ω_0 can be associated with the preferred rotation, or *twist*, of the tangent plane along the edge or pitch axis (Helfrich and Prost, 1988). At a subunit scale, this preference for mesoscopic twist derives form a energetic preferrence for locally skewed packing of molecules in the membrane (Zhang *et al.*, 2019), although a predictive understanding of the relationship between preferred pitch and structure of constituent chiral molecules is a notoriously complex and longstanding issue in and of itself; see Harris, Kamien, and Lubensky (1999).

Combining these and dividing by the number of subunits per ribbon $n = WL/a_0$ gives the excess energy

$$\epsilon_{\text{ex}}(W) = \frac{Y a_0}{1440} (C_{pp} C_{ee} - C_{ep}^2)^2 W^4 + \frac{B a_0}{2} [C_{pp}^2 + C_{ee}^2 + 2(C_{ep} - \Omega_0)^2], \tag{B6}$$

where, again, the curvature components in this expression are taken to correspond to the ribbon shape at the midline, and further assumed to be constant along the length of the ribbon. Reconsidering the comparison to the heuristic description in Eq. (59), inspection of the chirality-frustrated ribbon energy in Eq. (B6) shows that Gaussian curvature plays the role of frustration strength [i.e., $f \to K_{\text{G}} \simeq \det(C_{ij})$], while the shape parameter can be captured by the curvature tensor (i.e., $s \to C_{ij}$) with a preferred (tensorial) shape component that is nonzero only along the off-diagonal elements.

The shape equilibrium of the ribbon varies with width from the roots of the equations

$$\frac{\partial \epsilon_{\rm ex}}{\partial C_{ee}} = \frac{\partial \epsilon_{\rm ex}}{\partial C_{pp}} = \frac{\partial \epsilon_{\rm ex}}{\partial C_{ep}} = 0 \tag{B7}$$

for fixed W. The shape equilibrium is characterized by two branches. The first is the *helicoidal* branch,

$$\left.\begin{aligned} C_{ee} = C_{pp} &= 0 \\ C_{ep} + \tfrac{YW^4}{720B} C_{ep}^3 &= \Omega_0 \end{aligned}\right\} \quad \text{(helicoid).} \tag{B8}$$

This is the branch of equilibria discussed in the main text. Helicoid twist tends to its preferred value for narrow ribbons $C_{ep}(W \ll W_{\rm un}) \simeq \Omega_0$; and the helicoid unwinds in the wide limit as $C_{ep}(W \gg W_{\rm un}) \simeq \Omega_0 (W_{\rm un}/W)^{4/3}$, where the unwinding size $W_{\rm un} \equiv (720B/Y\Omega_0^2)^{1/4}$, defined in Eq. (63), characterizes the crossover width between the two regimes.

The second branch corresponds to a symmetry-breaking transition to a spiral ribbon shape,

$$\left.\begin{aligned} C_{ee} = C_{pp} &= \pm\sqrt{C_{ep}^2 - \Omega_0^2 (W_{\rm un}/W)^4} \\ C_{ep} &= \Omega_0/2 \end{aligned}\right\} \quad \text{(spiral ribbon).} \tag{B9}$$

Note that this branch exists only above a critical width $W_{\rm c} = \sqrt{2} W_{\rm un}$, for which C_{ee} and C_{pp} are real.

The critical value corresponds to an elastic instability. For $W < W_{\rm c}$ the helicoid branch is stable, and no spiral equilibrium exists. For $W \geq W_{\rm c}$, the helicoid branch becomes unstable, and the stable branches become the two degenerate spiral states, which differ by signs of C_{ee} and C_{pp}. The shape equilibria and excess energy of both branches are plotted in Figs. 17(b) and 17(c), respectively.

Notice that for wide ribbons the Gaussian curvature of both branches vanishes: for unstable helicoids $K_{\rm G} = -C_{ep}^2 (W \gg W_{\rm un}) \approx -\Omega_0^2 (W_{\rm un}/W)^{2/3}$, and for stable spiral ribbons $K_{\rm G}(W \geq W_c) = -\Omega_0^2 (W_{\rm un}/W)^4$. The difference in power law suggests a much more rapid expulsion with Gaussian curvature with increasing width of spirals. Notwithstanding the faster frustration escape of spiral ribbons, the accumulant analysis shown in the inset of Fig. 17(c) predicts that the maximum self-limiting size preempts the mechanical stability with $W_{\rm max} = 0.85 W_{\rm un} < W_{\rm c} = \sqrt{2} W_{\rm un}$. Hence, a generic prediction of this narrow-ribbon model is that self-limiting ribbons can only be helicoidal in shape (Ghafouri and Bruinsma, 2005; Armon *et al.*, 2014).

APPENDIX C: ASSEMBLY TIMESCALES AND KINETIC TRAPS

From Eq. (67) the initial nucleation timescale for the capsid model is

$$\tau_{\rm nuc}(\phi) \approx \frac{\exp\{-\beta\Omega[n_{\rm nuc}(\phi)]\}}{Z f_{\rm assem}[n_{\rm nuc}(\phi)]}, \tag{C1}$$

with $Z = \sqrt{\beta\tilde{\lambda}/\pi n_{\rm T}}(1+\Gamma^2)^{3/4}$ the Zeldovich factor that accounts for the time the system spends in the vicinity of the critical nucleus and $f_{\rm assem}(n_{\rm nuc}) \approx f_{\rm assem}$.

The median assembly time is then given by using Eq. (C1) to integrate the cumulative depletion of monomers as a function of time, approximately accounting for reversibility of the reaction as in Hagan and Elrad (2010), resulting in

$$\tau_{1/2} \approx 2^{n_{\rm nuc}(\Phi)-1}/[n_{\rm nuc}(\Phi) - 1] f_{\rm T} \frac{\tau_{\rm nuc}(\Phi)}{\Phi n_{\rm T}}, \tag{C2}$$

in which we made the approximation that the critical nucleus size remains constant over time ($n_{\rm nuc}[\phi_1(t)] \approx n_{\rm nuc}(\Phi) \equiv n_{\rm nuc}(\Phi)$).

REFERENCES

Achard, M. F., M. Kleman, Y. A. Nastishin, and H. T. Nguyen, 2005, Eur. Phys. J. E **16**, 37.

Agarwal, V., and B. Peters, 2014, *Solute Precipitate Nucleation: A Review of Theory and Simulation Advances* (John Wiley & Sons, New York), Chap. 3, pp. 97–160.

Aggeli, A., I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov, and N. Boden, 2001, Proc. Natl. Acad. Sci. U.S.A. **98**, 11857.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, 2002, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York).

Alder, B. J., and T. E. Wainwright, 1962, Phys. Rev. **127**, 359.

Armon, S., H. Aharoni, M. Moshe, and E. Sharon, 2014, Soft Matter **10**, 2733.

Auer, S., and D. Frenkel, 2001, Nature (London) **409**, 1020.

Baker, T. S., N. H. Olson, and S. D. Fuller, 2000, Microbiol. Mol. Biol. Rev. **64**, 237.

Bale, J. B., *et al.*, 2016, Science **353**, 389.

Bär, M., R. Grossmann, S. Heidenreich, and F. Peruani, 2020, Annu. Rev. Condens. Matter Phys. **11**, 441.

Basios, V., J. Lutsko, G. Nicolis, D. Maes, and C. Kirschhock, 2009, Microgravity Sci. Technol. **21**, 47.

Bausch, A. R., 2003, Science **299**, 1716.

Bechinger, C., R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, 2016, Rev. Mod. Phys. **88**, 045006.

Becker, R., and W. Döring, 1935, Ann. Phys. (Berlin) **416**, 719.

Ben-Shaul, A., I. Szleifer, and W. M. Gelbart, 1984, Proc. Natl. Acad. Sci. U.S.A. **81**, 4601.

Benson, E., A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen, and B. Högberg, 2015, Nature (London) **523**, 441.

Berger, B., P. W. Shor, L. Tuckerkellogg, and J. King, 1994, Proc. Natl. Acad. Sci. U.S.A. **91**, 7732.

Bergström, L. M., 2016, Curr. Opin. Colloid Interface Sci. **22**, 46.

Bernheim-Groswasser, A., R. Zana, and Y. Talmon, 2000, J. Phys. Chem. B **104**, 4005.

Binder, K., and D. Stauffer, 1976, Adv. Phys. **25**, 343.

Bobik, T. A., B. P. Lehman, and T. O. Yeates, 2015, Mol. Microbiol. **98**, 193.

Boles, M. A., M. Engel, and D. V. Talapin, 2016, Chem. Rev. **116**, 11220.

Bollinger, J. A., and M. J. Stevens, 2018, Soft Matter **14**, 1748.

Bowick, M. J., and L. Giomi, 2009, Adv. Phys. **58**, 449.

Briggs, J. A. G., J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H.-G. Krausslich, 2009, Proc. Natl. Acad. Sci. U.S.A. **106**, 11090.

Brillault, L., P. V. Jutras, N. Dashti, E. C. Thuenemann, G. Morgan, G. P. Lomonossoff, M. J. Landsberg, and F. Sainsbury, 2017, ACS Nano **11**, 3476.

Briscoe, J., and S. Small, 2015, Development (Cambridge, U.K.) **142**, 3996.

Broadhead, M. J., *et al.*, 2016, Sci. Rep. **6**, 24626.

Brocca, S., R. Grandori, S. Longhi, and V. Uversky, 2020, Int. J. Mol. Sci. **21**, 9045.

Brown, A. I., L. Kreplak, and A. D. Rutenberg, 2014, Soft Matter **10**, 8500.

Bruinsma, R. F., W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, 2003, Phys. Rev. Lett. **90**, 248101.

Bruinsma, R. F., and W. S. Klug, 2015, Annu. Rev. Condens. Matter Phys. **6**, 245.

Bucher, D., *et al.*, 2018, Nat. Commun. **9**, 1109.

Burlakov, V. M., N. Emptage, A. Goriely, and P. C. Bressloff, 2012, Phys. Rev. Lett. **108**, 028101.

Butterfield, G. L., *et al.*, 2017, Nature (London) **552**, 415.

Cameron, S., L. Kreplak, and A. D. Rutenberg, 2018, Soft Matter **14**, 4772.

Caplan, M. E., and C. J. Horowitz, 2017, Rev. Mod. Phys. **89**, 041002.

Cardinaux, F., A. Stradner, P. Schurtenberger, F. Sciortino, and E. Zaccarelli, 2007, Europhys. Lett. **77**, 48004.

Carlson, C. R., J. B. Asfaha, C. M. Ghent, C. J. Howard, N. Hartooni, and D. O. Morgan, 2020, Mol. Cell **80**, P1092.

Caspar, D. L. D., and A. Klug, 1962, Cold Spring Harbor Symp. Quant. Biol. **27**, 1.

Ceres, P., and A. Zlotnick, 2002a, Society **41**, 11525.

Ceres, P., and A. Zlotnick, 2002b, Biochemistry **41**, 11525.

Chen, T., Z. L. Zhang, and S. C. Glotzer, 2007, Proc. Natl. Acad. Sci. U.S.A. **104**, 717.

Chen, Z., N. Li, S. Li, M. Dharmarwardana, A. Schlimme, and J. J. Gassensmith, 2016, Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. **8**, 512.

Cheng, S., A. Aggarwal, and M. J. Stevens, 2012, Soft Matter **8**, 5666.

Chevreuil, M., L. Lecoq, S. Wang, L. Gargowitsch, N. Nhiri, E. Jacquet, T. Zinn, S. Fieulaine, S. Bressanelli, and G. Tresset, 2020, J. Phys. Chem. B **124**, 9987.

Cho, W.-K., J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and I. I. Cisse, 2018, Science **361**, 412.

Chong, S., *et al.*, 2018, Science **361**, eaar2555.

Chowdhury, C., S. Sinha, S. Chun, T. O. Yeates, and T. A. Bobik, 2014, Microbiol. Mol. Biol. Rev. **78**, 438.

Conchúir, B. O., and A. Zaccone, 2013, Phys. Rev. E **87**, 032310.

Cozzoli, P. D., T. Pellegrino, and L. Manna, 2006, Chem. Soc. Rev. **35**, 1195.

Crick, F. H. C., and J. D. Watson, 1956, Nature (London) **177**, 473.

Dang, X., H. Yi, M.-H. Ham, J. Qi, D. S. Yun, R. Ladewski, M. S. Strano, P. T. Hammond, and A. M. Belcher, 2011, Nat. Nanotechnol. **6**, 377.

Debye, P., 1949, Ann. N.Y. Acad. Sci. **51**, 575.

Delalande, L., I. B. Tsvetkova, C. Zeng, K. Bond, M. F. Jarrold, and B. Dragnea, 2016, Nanoscale **8**, 16221.

de Pablo, J. J., *et al.*, 2019, npj Comput. Mater. **5**, 41.

Desai, A., and T. J. Mitchison, 1997, Annu. Rev. Cell Dev. Biol. **13**, 83.

De Yoreo, J. J., and P. G. Vekilov, 2003, Rev. Mineral. Geochem. **54**, 57.

Dharmavaram, S., S. B. She, G. Lázaro, M. F. Hagan, and R. Bruinsma, 2019, PLoS Comput. Biol. **15**, e1006602.

DiDonna, B. A., and R. D. Kamien, 2003, Phys. Rev. E **68**, 041703.

Dill, K. A., and P. J. Flory, 1980, Proc. Natl. Acad. Sci. U.S.A. **77**, 3115.

Dinsmore, A., P. Dubin, and G. Grason, 2011, J. Phys. Chem. B **115**, 7173.

Doostmohammadi, A., J. Ignes-Mullol, J. M. Yeomans, and F. Sagues, 2018, Nat. Commun. **9**, 3246.

Dufresne, E. R., H. Noh, V. Saranathan, S. G. J. Mochrie, H. Cao, and R. O. Prum, 2009, Soft Matter **5**, 1792.

Elrad, O. M., and M. F. Hagan, 2008, Nano Lett. **8**, 3850.

Elrad, O. M., and M. F. Hagan, 2010, Phys. Biol. **7**, 045003.

Endres, D., and A. Zlotnick, 2002, Biophys. J. **83**, 1217.

Fai, T. G., L. Mohapatra, P. Kar, J. Kondev, A. Amir, F. Jülicher, and N. Barkai, 2019, eLife **8**, e42599.

Fejer, S. N., D. Chakrabarti, and D. J. Wales, 2010, ACS Nano **4**, 219.

Fernández de Castro, I., R. Tenorio, and C. Risco, 2020, in *Reference Module in Life Sciences*, edited by B. D. Roitberg (Elsevier, New York).

Fisher, M. E., 1984, J. Stat. Phys. **34**, 667.

Foderà, V., A. Zaccone, M. Lattuada, and A. M. Donald, 2013, Phys. Rev. Lett. **111**, 108105.

Foret, L., and P. Sens, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 14763.

Fortini, A., E. Sanz, and M. Dijkstra, 2008, Phys. Rev. E **78**, 041402.

Fox, J. M., G. J. Wang, J. A. Speir, N. H. Olson, J. E. Johnson, T. S. Baker, and M. J. Young, 1998, Virology **244**, 212.

Fraenkel-Conrat, H., and R. C. Williams, 1955, Proc. Natl. Acad. Sci. U.S.A. **41**, 690.

Fratzl, P., 2003, Curr. Opin. Colloid Interface Sci. **8**, 32.

Garmann, R. F., M. Comas-Garcia, A. Gopal, C. M. Knobler, and W. M. Gelbart, 2014, J. Mol. Biol. **426**, 1050.

Garmann, R. F., M. Comas-Garcia, C. M. Knobler, and W. M. Gelbart, 2016, Acc. Chem. Res. **49**, 48.

Gelbart, W., A. Ben-Shaul, and D. Roux, 1994, *Micelles, Membranes, Microemulsions and Monolayers* (Springer, New York).

Geng, Y., P. Dalhaimer, S. Cai, R. Tsai, M. Tewari, T. Minko, and D. E. Discher, 2007, Nat. Nanotechnol. **2**, 249.

Ghafouri, R., and R. Bruinsma, 2005, Phys. Rev. Lett. **94**, 138101.

Giani, M., W. K. D. Otter, W. J. Briels, M. Giani, W. K. D. Otter, and W. J. Briels, 2017, J. Chem. Phys. **146**, 155102.

Gibaud, T., *et al.*, 2012, Nature (London) **481**, 348.

Gliko, O., N. Neumaier, W. Pan, I. Haase, M. Fischer, A. Bacher, S. Weinkauf, and P. G. Vekilov, 2005, J. Am. Chem. Soc. **127**, 3433.

Glotzer, S. C., and M. J. Solomon, 2007, Nat. Mater. **6**, 557.

Grant, J., and R. L. Jack, 2012, Phys. Rev. E **85**, 021112.

Grant, J., R. L. Jack, and S. Whitelam, 2011, J. Chem. Phys. **135**, 214505.

Grason, G. M., 2012, Phys. Rev. E **85**, 031603.

Grason, G. M., 2015, Rev. Mod. Phys. **87**, 401.

Grason, G. M., 2016, J. Chem. Phys. **145**, 110901.

Grason, G. M., 2017, Nat. Phys. **13**, 1149.

Grason, G. M., 2020, Soft Matter **16**, 1102.

Grason, G. M., and R. F. Bruinsma, 2007, Phys. Rev. Lett. **99**, 098101.

Green, L. N., H. K. K. Subramanian, V. Mardanlou, J. Kim, R. F. Hariadi, and E. Franco, 2019, Nat. Chem. **11**, 510.

Groenewold, J., and W. K. Kegel, 2001, J. Phys. Chem. B **105**, 11702.

Grossman, D., E. Sharon, and H. Diamant, 2016, Phys. Rev. Lett. **116**, 258105.

Grover, M. A., D. J. Griffin, and X. Tang, 2019, IFAC-PapersOnLine **52**, 1.

Guseva, S., S. Milles, M. R. Jensen, N. Salvi, J.-P. Kleman, D. Maurin, R. W. H. Ruigrok, and M. Blackledge, 2020, Sci. Adv. **6**, eaaz7095.

Haddad, A., H. Aharoni, E. Sharon, A. G. Shtukenberg, B. Kahr, and E. Efrati, 2019, Soft Matter **15**, 116.

Hagan, M. F., 2008, Phys. Rev. E **77**, 051904.

Hagan, M. F., 2014, Adv. Chem. Phys. **155**, 1.

Hagan, M. F., and D. Chandler, 2006, Biophys. J. **91**, 42.

Hagan, M. F., and O. M. Elrad, 2010, Biophys. J. **98**, 1065.

Hagan, M. F., O. M. Elrad, and R. L. Jack, 2011, J. Chem. Phys. **135**, 104115.

Hagan, M. F., and R. Zandi, 2016, Curr. Opin. Virol. **18**, 36.

Halatek, J., and E. Frey, 2018, Nat. Phys. **14**, 507.

Hall, D. M., I. R. Bruss, J. R. Barone, and G. M. Grason, 2016, Nat. Mater. **15**, 727.

Hall, D. M., and G. M. Grason, 2017, Interface Focus **7**, 20160140.

Halperin, A., M. Tirrell, and T. P. Lodge, 1992, in *Macromolecules: Synthesis, Order and Advanced Properties*, Advances in Polymer Science Vol. 100, edited by K. A. Armitstead *et al.* (Springer-Verlag, Berlin), pp. 31–71.

Hamley, I. W., 2003, Angew. Chem., Int. Ed. Engl. **42**, 1692.

Harris, A. B., R. D. Kamien, and T. C. Lubensky, 1999, Rev. Mod. Phys. **71**, 1745.

Haselwandter, C. A., M. Calamai, M. Kardar, A. Triller, and R. A. da Silveira, 2011, Phys. Rev. Lett. **106**, 238104.

Haselwandter, C. A., M. Kardar, A. Triller, and R. A. da Silveira, 2015, Phys. Rev. E **92**, 032705.

Helfrich, W., 1986, J. Phys. (Paris) **47**, 321.

Helfrich, W., and J. Prost, 1988, Phys. Rev. A **38**, 3065.

Hemmat, M., and D. J. Odde, 2021, Ann. Biomed. Eng. (in press), https://doi.org/10.1007/s10439-020-02715-6.

Hiemenz, P. C., and T. P. Lodge, 2007, *Polymer Chemistry*, 2nd ed. (CRC Press, Baton Rouge).

Hnisz, D., K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, 2017, Cell **169**, 13.

Hormoz, S., and M. P. Brenner, 2011, Proc. Natl. Acad. Sci. U.S.A. **108**, 5193.

Hough, L. E., *et al.*, 2009, Science **325**, 456.

Iancu, C. V., H. J. Ding, D. M. Morris, D. P. Dias, A. D. Gonzales, A. Martino, and G. J. Jensen, 2007, J. Mol. Biol. **372**, 764.

Irvine, W. T. M., V. Vitelli, and P. M. Chaikin, 2010, Nature (London) **468**, 947.

Israelachvili, J., 2011, *Intermolecular and Surface Forces* (Elsevier, New York).

Israelachvili, J. N., D. J. Mitchell, and B. W. Ninham, 1976, J. Chem. Soc. Faraday Trans. 2 **72**, 1525.

Jack, R. L., M. F. Hagan, and D. Chandler, 2007, Phys. Rev. E **76**, 021119.

Jacobs, W. M., and D. Frenkel, 2016, J. Am. Chem. Soc. **138**, 2457.

Jacobs, W. M., A. Reinhardt, and D. Frenkel, 2015, Proc. Natl. Acad. Sci. U.S.A. **112**, 6313.

Jacobson, L. C., W. Hujo, and V. Molinero, 2010, J. Am. Chem. Soc. **132**, 11806.

Jain, S., and F. S. Bates, 2003, Science **300**, 460.

Johnson, J. E., and J. A. Speir, 1997, J. Mol. Biol. **269**, 665.

Jones, M. R., N. C. Seeman, and C. A. Mirkin, 2015, Science **347**, 1260901.

Joswiak, M. N., N. Duff, M. F. Doherty, and B. Peters, 2013, J. Phys. Chem. Lett. **4**, 4267.

Judd, J., M. L. Ho, A. Tiwari, E. J. Gomez, C. Dempsey, K. Van Vliet, O. A. Igoshin, J. J. Silberg, M. Agbandje-McKenna, and J. Suh, 2014, ACS Nano **8**, 4740.

Kang, L., and T. C. Lubensky, 2017, Proc. Natl. Acad. Sci. U.S.A. **114**, E19.

Ke, Y., L. L. Ong, W. M. Shih, and P. Yin, 2012, Science **338**, 1177.

Ke, Y., L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih, and P. Yin, 2014, Nat. Chem. **6**, 994.

Kerfeld, C. A., S. Heinhorst, and G. C. Cannon, 2010, Microbiology **64**, 391.

Kerfeld, C. A., and M. R. Melnicki, 2016, Curr. Opin. Plant Biol. **31**, 66.

Kieser, Q., R. S. Noyce, M. Shenouda, Y. J. Lin, and D. H. Evans, 2020, PLoS One **15**, e0228028.

King, N. P., J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates, and D. Baker, 2014, Nature (London) **510**, 103.

Kirchhausen, T., D. Owen, and S. C. Harrison, 2014, Cold Spring Harbor Perspect. Biol. **6**, a016725.

Kléman, M., 1989, Adv. Phys. **38**, 605.

Klok, H.-A., and S. Lecommandoux, 2001, Adv. Mater. **13**, 1217.

Klotsa, D., and R. L. Jack, 2011, Soft Matter **7**, 6294.

Klotsa, D., and R. L. Jack, 2013, J. Chem. Phys. **138**, 094502.

Knott, B. C., V. Molinero, M. F. Doherty, and B. Peters, 2012, J. Am. Chem. Soc. **134**, 19544.

Lai, Y. T., E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson, and T. O. Yeates, 2014, Nat. Chem. **6**, 1065.

Lázaro, G. R., B. Dragnea, and M. F. Hagan, 2018, Soft Matter **14**, 5728.

Lázaro, G. R., S. Mukhopadhyay, and M. F. Hagan, 2018, Biophys. J. **114**, 619.

Leibler, L., H. Orland, and J. C. Wheeler, 1983, J. Chem. Phys. **79**, 3550.

Lenz, M., and T. A. Witten, 2017, Nat. Phys. **13**, 1100.

LeRoy, H., 2018, master's thesis (Université Paris–Saclay).

Li, C., A. G. Shtukenberg, L. Vogt-Maranto, E. Efrati, P. Raiteri, J. D. Gale, A. L. Rohl, and B. Kahr, 2020, J. Phys. Chem. C **124**, 15616.

Li, F., D. P. Josephson, and A. Stein, 2011, Angew. Chem., Int. Ed. Engl. **50**, 360.

Li, S., P. Roy, A. Travesset, and R. Zandi, 2018, Proc. Natl. Acad. Sci. U.S.A. **115**, 10971.

Li, S., R. Zandi, A. Travesset, and G. M. Grason, 2019, Phys. Rev. Lett. **123**, 145501.

Lisman, J., and S. Raghavachari, 2006, Sci. STKE **2006**, re11.

Lisman, J., and S. Raghavachari, 2015, Brain Res. **1621**, 51.

Liu, K. K. L., M. F. Hagan, and J. E. Lisman, 2017, Phil. Trans. R. Soc. B **372**, 20160328.

Loeffler, T. D., D. E. Henderson, B. Chen, T. D. Loeffler, D. E. Henderson, and B. Chen, 2012, J. Chem. Phys. **137**, 194304.

Luque, A., and D. Reguera, 2010, Biophys. J. **98**, 2993.

Lutomski, C., N. A. Lyktey, E. E. Pierson, Z. Zhao, A. Zlotnick, and M. F. Jarrold, 2018, J. Am. Chem. Soc. **140**, 5784.

Malyutin, A. G., *et al.*, 2015, Chem. Mater. **27**, 327.

Manoharan, V. N., 2015, Science **349**, 1253751.

Marchetti, M., J. Joanny, S. Ramaswamy, T. Liverpool, J. Prost, M. Rao, and R. A. Simha, 2013, Rev. Mod. Phys. **85**, 1143.

Marsland, R., and J. L. England, 2018, Phys. Rev. E **98**, 022411.

Mateu, M. G., 2013, Arch. Biochem. Biophys. **531**, 65.

Matsumoto, E. A., G. P. Alexander, and R. D. Kamien, 2009, Phys. Rev. Lett. **103**, 257804.

May, E. R., A. Aggarwal, W. S. Klug, and C. L. Brooks, 2011, Biophys. J. **100**, L59.

May, E. R., and C. L. Brooks, 2011, Phys. Rev. Lett. **106**, 188101.

May, S., and A. Ben-Shaul, 2001, J. Phys. Chem. B **105**, 630.

Mayer, J. E., and W. W. Wood, 1965, J. Chem. Phys. **42**, 4268.

Mbanga, B. L., G. M. Grason, and C. D. Santangelo, 2012, Phys. Rev. Lett. **108**, 017801.

McGraw, R., and A. Laaksonen, 1997, J. Chem. Phys. **106**, 5284.

McPhedran, R. C., and A. R. Parker, 2015, Phys. Today **68**, No. 6, 32.

Meek, K. M., 2009, Biophys. Rev. Lett. **1**, 83.

Meiri, S., and E. Efrati, 2021, arXiv:2101.09816.

Mendoza, C. I., and D. Reguera, 2020, eLife **9**, e52525.

Meng, G., J. Paulose, D. R. Nelson, and V. N. Manoharan, 2014, Science **343**, 634.

Mettlen, M., P. H. Chen, S. Srinivasan, G. Danuser, and S. L. Schmid, 2018, Annu. Rev. Biochem. **87**, 871.

Miermans, C. A., R. P. T. Kusters, C. C. Hoogenraad, and C. Storm, 2017, PLoS One **12**, e0170113.

Minten, I. J., V. I. Claessen, K. Blank, A. E. Rowan, R. J. M. Nolte, and J. J. L. M. Cornelissen, 2011, Chem. Sci. **2**, 358.

Missel, P. J., N. A. Mazer, G. B. Benedek, C. Y. Young, and M. C. Carey, 1980, J. Phys. Chem. **84**, 1044.

Mohammed, A. M., and R. Schulman, 2013, Nano Lett. **13**, 4006.

Mohapatra, L., B. L. Goode, P. Jelenkovic, R. Phillips, and J. Kondev, 2016, Annu. Rev. Biophys. **45**, 85.

Morphew, D., and D. Chakrabarti, 2017, Curr. Opin. Colloid Interface Sci. **30**, 70.

Morphew, D., J. Shaw, C. Avins, and D. Chakrabarti, 2018, ACS Nano **12**, 2355.

Morton, V. L., E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, and P. G. Stockley, 2010, J. Mol. Biol. **401**, 298.

Mosayebi, M., D. K. Shoemark, J. M. Fletcher, R. B. Sessions, N. Linden, D. N. Woolfson, and T. B. Liverpool, 2017, Proc. Natl. Acad. Sci. U.S.A. **114**, 9014.

Murugan, A., J. Zou, and M. P. Brenner, 2015, Nat. Commun. **6**, 6203.

Nagarajan, R., 2002, Langmuir **18**, 31.

Nagarajan, R., and E. Ruckenstein, 1991, Langmuir **7**, 2934.

Namba, K., and F. Vonderviszt, 1997, Q. Rev. Biophys. **30**, 1.

Needleman, D., and Z. Dogic, 2017, Nat. Rev. Mater. **2**, 17048.

Nelson, D., and L. Peliti, 1987, J. Phys. (Paris) **48**, 1085.

Neville, A. C., 1993, *Biology of Fibrous Composites* (Cambridge University Press, Cambridge, England).

Nguyen, H. D., V. S. Reddy, and C. L. Brooks, 2007, Nano Lett. **7**, 338.

Nguyen, H. D., V. S. Reddy, and C. L. Brooks, 2009, J. Am. Chem. Soc. **131**, 2606.

Nguyen, M., and S. Vaikuntanathan, 2016, Proc. Natl. Acad. Sci. U.S.A. **113**, 14231.

Nguyen, T. D., B. A. Schultz, N. A. Kotov, and S. C. Glotzer, 2015, Proc. Natl. Acad. Sci. U.S.A. **112**, E3161.

Nicolis, G., and C. Nicolis, 2003, Physica (Amsterdam) **323A**, 139.

Nikolic, J., R. Le Bars, Z. Lama, N. Scrima, C. Lagaudriere-Gesbert, Y. Gaudin, and D. Blondel, 2017, Nat. Commun. **8**, 58.

Niv, I., and E. Efrati, 2018, Soft Matter **14**, 424.

Nogales, E., 2000, Annu. Rev. Biochem. **69**, 277.

Nott, T. J., *et al.*, 2015, Mol. Cell **57**, 936.

O'Brien, M. N., M. R. Jones, and C. A. Mirkin, 2016, Proc. Natl. Acad. Sci. U.S.A. **113**, 11717.

Oda, R., I. Huc, M. Schmutz, S. J. Candau, and F. C. MacKintosh, 1999, Nature (London) **399**6736, 566.

Oltra, N. S., P. Nair, and D. E. Discher, 2014, Annu. Rev. Chem. Biomol. Eng. **5**, 281.

Ong, L. L., *et al.*, 2017, Nature (London) **552**, 72.

Oosawa, F., and S. Asakura, 1975, *Thermodynamics of the Polymerization of Protein* (Academic Press, London).

Ottani, V., D. Martini, M. Franchi, A. Ruggeri, and M. Raspanti, 2002, Micron **33**, 587.

Oxtoby, D. W., 1992, J. Phys. Condens. Matter **4**, 7627.

Pan, A. C., and D. Chandler, 2004, J. Phys. Chem. B **108**, 19681.

Panahandeh, S., S. Li, L. Marichal, R. Leite Rubim, G. Tresset, and R. Zandi, 2020, ACS Nano **14**, 3170.

Panahandeh, S., S. Li, and R. Zandi, 2018, Nanoscale **10**, 22802.

Park, H., *et al.*, 2016, Nat. Mater. **15**, 211.

Perkett, M. R., D. T. Mirijanian, and M. F. Hagan, 2016, J. Chem. Phys. **145**, 035101.

Perlmutter, J. D., and M. F. Hagan, 2015, Annu. Rev. Phys. Chem. **66**, 217.

Perlmutter, J. D., M. R. Perkett, and M. F. Hagan, 2014, J. Mol. Biol. **426**, 3148.

Peters, B., 2009, J. Chem. Phys. **131**, 244103.

Pfeifer, F., 2012, Nat. Rev. Microbiol. **10**, 705.

Pierson, E. E., D. Z. Keifer, A. A. Kukreja, J. C. Wang, A. Zlotnick, and M. F. Jarrold, 2016, J. Mol. Biol. **428**, 292.

Pierson, E. E., D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. Y. Wang, A. Zlotnick, and M. F. Jarrold, 2014, J. Am. Chem. Soc. **136**, 3536.

Pineros, W. D., B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, 2018, J. Chem. Phys. **148**, 104509.

Pollard, T. D., 2016, Cold Spring Harbor Perspect. Biol. **8**, a018226.

Popp, D., and R. C. Robinson, 2012, Cytoskeleton **69**, 71.

Porte, G., Y. Poggi, J. Appell, and G. Maret, 1984, J. Phys. Chem. **88**, 5713.

Prasad, B. V. V., and M. F. Schmid, 2012, in *Viral Molecular Machines*, edited by M. G. Rossmann and V. B. Rao (Springer, Boston), pp. 17–47.

Prum, R. O., E. R. Dufresne, T. Quinn, and K. Waters, 2009, J. R. Soc. Interface **6**, S253.

Rae, B. D., B. M. Long, M. R. Badger, and G. D. Price, 2013, Microbiol. Mol. Biol. Rev. **77**, 357.

Rapaport, D., 2008, Phys. Rev. Lett. **101**, 186101.

Reguera, D., R. K. Bowles, Y. Djikaev, and H. Reiss, 2003, J. Chem. Phys. **118**, 340.

Reguera, D., J. Hernández-Rojas, and J. M. Gomez Llorente, 2019, Soft Matter **15**, 7166.

Roos, W. H., R. Bruinsma, and G. J. L. Wuite, 2010, Nat. Phys. **6**, 733.

Roos, W. H., and G. L. Wuite, 2009, Adv. Mater. **21**, 1187.

Rothemund, P. W. K., A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson, and E. Winfree, 2004, J. Am. Chem. Soc. **126**, 16344.

Rother, M., M. G. Nussbaumer, K. Renggli, and N. Bruns, 2016, Chem. Soc. Rev. **45**, 6213.

Royle, S. J., 2012, J. Cell Sci. **125**, 19.

Ruan, L., J. A. Hadden, and A. Zlotnick, 2018, J. Virol. **92**, e01082-18.

Russel, W. B., D. A. Saville, and W. R. Schowalter, 1989, *Colloidal Dispersions* (Cambridge University Press, Cambridge, England).

Sabari, B. R., *et al.*, 2018, Science **361**, eaar3958.

Sacanna, S., and D. J. Pine, 2011, Curr. Opin. Colloid Interface Sci. **16**, 96.

Sadoc, J.-F., and R. Mosseri, 2006, *Geometrical Frustration* (Cambridge University Press, Cambridge, England).

Safran, S., 1994, *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes* (Addison-Wesley, Reading, MA).

Sakhardande, R., S. Stanojeviea, A. Baskaran, A. Baskaran, M. F. Hagan, and B. Chakraborty, 2017, Phys. Rev. E **96**, 012704.

Saranathan, V., J. D. Forster, H. Noh, S.-F. Liew, S. G. J. Mochrie, H. Cao, E. R. Dufresne, and R. O. Prum, 2012, J. R. Soc. Interface **9**, 2563.

Savastano, A., A. Ibanez de Opakua, M. Rankovic, and M. Zweckstetter, 2020, Nat. Commun. **11**, 6041.

Schmid, M. F., A. M. Paredes, H. A. Khant, F. Soyer, H. C. Aldrich, W. Chiu, and J. M. Shively, 2006, J. Mol. Biol. **364**, 526.

Schneider, S., and G. Gompper, 2005, Europhys. Lett. **70**, 136.

Schoelz, J. E., and S. Leisner, 2017, Front. Plant Sci. **8**, 1832.

Schubert, R., A. Meyer, D. Baitan, K. Dierks, M. Perbandt, and C. Betzel, 2017, Cryst. Growth Des. **17**, 954.

Sciortino, F., S. Mossa, E. Zaccarelli, and P. Tartaglia, 2004, Phys. Rev. Lett. **93**, 055701.

Sciortino, F., P. Tartaglia, and E. Zaccarelli, 2005, J. Phys. Chem. B **109**, 21942.

Sear, R. P., 2007, J. Phys. Condens. Matter **19**, 033101.

Sear, R. P., 2009, J. Chem. Phys. **131**, 074702.

Sear, R. P., and W. M. Gelbart, 1999, J. Chem. Phys. **110**, 4582.

Sedgwick, H., S. U. Egelhaaf, and W. C. K. Poon, 2004, J. Phys. Condens. Matter **16**, S4913.

Seifert, U., 2008, Eur. Phys. J. B **64**, 423.

Seifert, U., 2012, Rep. Prog. Phys. **75**, 126001.

Selinger, J. V., M. S. Spector, and J. M. Schnur, 2001, J. Phys. Chem. B **105**, 7157.

Selinger, R. L. B., J. V. Selinger, A. P. Malanoski, and J. M. Schnur, 2004, Phys. Rev. Lett. **93**, 158103.

Selzer, L., S. P. Katen, and A. Zlotnick, 2014, Biochemistry **53**, 5496.

Seul, M., and D. Andelman, 1995, Science **267**5197, 476.

Seung, H. S., and D. R. Nelson, 1988, Phys. Rev. A **38**, 1005.

Shaebani, M. R., A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger, 2020, Nat. Rev. Phys. **2**, 181.

Sharma, P., A. Ward, T. Gibaud, M. F. Hagan, and Z. Dogic, 2014, Nature (London) **513**, 77.

Shomar, A., L. Geyrhofer, N. E. Ziv, and N. Brenner, 2017, PLoS Comput. Biol. **13**, e1005668.

Siggia, E. D., 1979, Phys. Rev. A **20**, 595.

Sigl, C., *et al.*, 2021, Nat. Mater. (in press).

Statt, A., P. Virnau, and K. Binder, 2015, Phys. Rev. Lett. **114**, 026101.

Steinmetz, N. F., C.-F. Cho, A. Ablack, J. D. Lewis, and M. Manchester, 2011, Nanomedicine **6**, 351.

Stockley, P. G., N. A. Ranson, and R. Twarock, 2013, Future Virol. **8**, 531.

Stradner, A., H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger, 2004, Nature (London) **432**, 492.

Stupp, S. I., and L. C. Palmer, 2014, Chem. Mater. **26**, 507.

Su, Z., R. Zhang, X.-Y. Yan, Q.-Y. Guo, J. Huang, W. Shan, Y. Liu, T. Liu, M. Huang, and S. Z. Cheng, 2020, Prog. Polym. Sci. **103**, 101230.

Sun, J., *et al.*, 2007, Proc. Natl. Acad. Sci. U.S.A. **104**, 1354.

Sutter, M., D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, and N. Ban, 2008, Nat. Struct. Mol. Biol. **15**, 939.

Tanaka, S., C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, and T. O. Yeates, 2008, Science **319**, 1083.

Tanford, C., 1974, J. Phys. Chem. **78**, 2469.

Tang, A.-H., H. Chen, T. P. Li, S. R. Metzbower, H. D. MacGillavry, and T. A. Blanpied, 2016, Nature (London) **536**, 210.

Tang, X., B. Rupp, Y. Yang, T. D. Edwards, M. A. Grover, and M. A. Bevan, 2016, ACS Nano **10**, 6791.

Tang, X., J. Zhang, M. A. Bevan, and M. A. Grover, 2017, J. Process Control **60**, 141.

ten Wolde, P. R., and D. Frenkel, 1997, Science **277**, 1975.

Terzi, M., H. LeRoy, and M. Lenz, 2020 (to be published).

Thompson, S. M., K. E. Gubbins, J. P. R. B. Walton, R. A. R. Chantry, and J. S. Rowlinson, 1984, J. Chem. Phys. **81**, 530.

Tian, C., X. Li, Z. Liu, W. Jiang, G. Wang, and C. Mao, 2014, Angew. Chem., Int. Ed. Engl. **53**, 8041.

Tong, D., and G. A. Voth, 2020, Biophys. J. **118**, 2938.

Tsvetkova, I., C. Chen, S. Rana, C. C. Kao, V. M. Rotello, and B. Dragnea, 2012, Soft Matter **8**, 4571.

Turing, A. M., 1952, Phil. Trans. R. Soc. B **237**, 37.

Turner, M. S., R. W. Briehl, F. A. Ferrone, and R. Josephs, 2003, Phys. Rev. Lett. **90**, 128103.

Twarock, R., 2004, J. Theor. Biol. **226**, 477.

Twarock, R., R. J. Bingham, E. C. Dykeman, and P. G. Stockley, 2018, Curr. Opin. Virol. **31**, 74.

Twarock, R., and A. Luque, 2019, Nat. Commun. **10**, 4414.

Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. van Duijn, and J. R. Heck Albert, 2011, Nat. Chem. **3**, 126.

Uetrecht, C., C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 9216.

Vannimenus, J., and G. Toulouse, 1977, J. Phys. C **10**, L537.

van Schooneveld, M. M., V. W. de Villeneuve, R. P. Dullens, D. G. Aarts, M. E. Leunissen, and W. K. Kegel, 2009, J. Phys. Chem. B **113**, 4560.

Veesler, S., E. Revalor, O. Bottini, and C. Hoff, 2006, Org. Process Res. Dev. **10**, 841.

Vitelli, V., and A. M. Turner, 2004, Phys. Rev. Lett. **93**, 215301.

Wang, C. J., S. Mukhopadhyay, and A. Zlotnick, 2018, Viruses **10**, 25.

Wang, J., V. Rayaprolu, S. Mukhopadhyay, and A. Zlotnick, 2015, ACS Nano **9**, 8898.

Weber, C. A., D. Zwicker, F. Julicher, and C. F. Lee, 2019, Rep. Prog. Phys. **82**, 064601.

Weisel, J. W., 2004, Biophys. Chem. **112**, 267.

Wess, T. J., 2008, in *Collagen: Structure and Mechanics*, edited by P. Fratzl (Springer, Boston), Chap. 3, pp. 49–80.

Whitelam, S., 2010, Phys. Rev. Lett. **105**, 088102.

Whitelam, S., E. H. Feng, M. F. Hagan, and P. L. Geissler, 2009, Soft Matter **5**, 1251.

Whitelam, S., and R. L. Jack, 2015, Annu. Rev. Phys. Chem. **66**, 143.

Whitesides, G. M., and B. Grzybowski, 2002, Science **295**, 2418.

Wilber, A. W., J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, 2007, J. Chem. Phys. **127**, 085106.

Wilber, A. W., J. P. K. Doye, A. A. Louis, and A. C. F. Lewis, 2009, J. Chem. Phys. **131**, 175102.

Wingfield, P. T., S. J. Stahl, R. W. Williams, and A. C. Steven, 1995, Biochemistry **34**, 4919.

Woltornist, S. J., J.-M. Y. Carrillo, T. O. Xu, A. V. Dobrynin, and D. H. Adamson, 2015, Macromolecules **48**, 687.

Woltornist, S. J., D. Varghese, D. Massucci, Z. Cao, A. V. Dobrynin, and D. H. Adamson, 2017, Adv. Mater. **29**, 1604947.

Wynne, S. A., R. A. Crowther, and A. G. W. Leslie, 1999, Mol. Cell **3**, 771.

Yang, Y., R. B. Meyer, and M. F. Hagan, 2010, Phys. Rev. Lett. **104**, 258102.

Yin, Y., and A. P. Alivisatos, 2005, Nature (London) **437**, 664.

Young, M., W. Debbie, M. Uchida, and T. Douglas, 2008, Annu. Rev. Phytopathol. **46**, 361.

Yu, N., A. Ghosh, and M. F. Hagan, 2016, Soft Matter **12**, 8990.

Zaccarelli, E., 2007, J. Phys. Condens. Matter **19**, 323101.

Zandi, R., B. Dragnea, A. Travesset, and R. Podgornik, 2020, Phys. Rep. **847**, 1.

Zandi, R., D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, 2004, *Proc. Natl. Acad. Sci. U.S.A.* 101, 15556.

Zandi, R., P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, 2006, Biophys. J. **90**, 1939.

Zaslavsky, B. Y., L. A. Ferreira, A. L. Darling, and V. N. Uversky, 2018, Int. J. Biol. Macromol. **117**, 1224.

Zeravcic, Z., V. N. Manoharan, and M. P. Brenner, 2014, Proc. Natl. Acad. Sci. U.S.A. **111**, 15918.

Zeravcic, Z., V. N. Manoharan, and M. P. Brenner, 2017, Rev. Mod. Phys. **89**, 031001.

Zhang, L., and A. Eisenberg, 1996, J. Am. Chem. Soc. **118**, 3168.

Zhang, M., D. Grossman, D. Danino, and E. Sharon, 2019, Nat. Commun. **10**, 3565.

Zhong, W., D. J. Schwab, and A. Murugan, 2017, J. Stat. Phys. **167**, 806.

Zhou, J., P. Kondylis, D. G. Haywood, Z. D. Harms, L. S. Lee, A. Zlotnick, and S. C. Jacobson, 2018, Anal. Chem. **90**, 7267.

Zhou, K., L. Li, Z. Tan, A. Zlotnick, and S. C. Jacobson, 2011, J. Am. Chem. Soc. **133**, 1618.

Zhuang, Y., and P. Charbonneau, 2016, J. Phys. Chem. B **120**, 7775.

Zimmermann, N. E. R., B. Vorselaars, D. Quigley, and B. Peters, 2015, J. Am. Chem. Soc. **137**, 13352.

Ziserman, L., A. Mor, D. Harries, and D. Danino, 2011, Phys. Rev. Lett. **106**, 238105.

Zlotnick, A., 2003, Virology **315**, 269.

Zlotnick, A., R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, 2000, Virology **277**, 450.

Zlotnick, A., J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, 1999, Biochemistry **38**, 14644.

Zlotnick, A., and S. Mukhopadhyay, 2011, Trends Microbiol. **19**, 14.

Zwicker, D., M. Decker, S. Jaensch, A. A. Hyman, and F. Julicher, 2014, Proc. Natl. Acad. Sci. U.S.A. **111**, E2636.

Zwicker, D., A. A. Hyman, and F. Julicher, 2015, Phys. Rev. E **92**, 012317.