Mechanisms of scaffold-mediated microcompartment assembly and size-control

Farzaneh Mohajerani¹, Evan Sayer¹, Christopher Neil², Koe Inlow², Michael F. Hagan¹

¹ *Martin Fisher School of Physics, Brandeis University, Waltham, Massachusetts, USA*

² *Department of Biochemistry, Brandeis University, Waltham, Massachusetts, USA*

This article describes a theoretical and computational study of the dynamical assembly of a protein shell around a complex consisting of many cargo molecules and long flexible scaffold molecules. Our study is motivated by bacterial microcompartments, which are proteinaceous organelles that assemble around a condensed droplet of enzymes and reactants. As in many examples of cytoplasmic liquid-liquid phase separation, condensation of the microcompartment interior cargo is driven by long flexible scaffold proteins that have weak multivalent interactions with the cargo. We describe a minimal model for the thermodynamics and dynamics of assembly of a protein shell around cargo and scaffold molecules, with scaffold-mediated cargo coalescence and encapsulation. Our results predict that the shell size, amount of encapsulated cargo, and assembly pathways depend sensitively on properties of the scaffold, including its length and valency of scaffold-cargo interactions. Moreover, the ability of self-assembling protein shells to change their size to accommodate scaffold molecules of different lengths depends crucially on whether the spontaneous curvature radius of the protein shell is smaller or larger than a characteristic elastic length scale of the shell. Beyond natural microcompartments, these results have important implications for synthetic biology efforts to target new molecules for encapsulation by microcompartments or viral shells. More broadly, the results elucidate how cells exploit coupling between self-assembly and liquid-liquid phase separation to organize their interiors.

While the eukaryotic cytoplasm is organized by lipid-encased organelles, it is now clear that cells from all kingdoms of life employ other modes of compartmentalization, such as liquid-liquid phase separation [1–4] and proteinaceous organelles [5–7]. For example, bacterial microcompartments are organelles that form by assembling a protein shell around a dense complex of enzymes and reactants in certain metabolic pathways [8–15]. The best-studied type of microcompartment is the carboxysome, which assembles around rubisco and carbonic anhydrase to enable carbon fixation in cyanobacteria [8, 9, 16, 17].

Other protein-shelled compartments are found in bacteria and archaea (e.g. encapsulins [18, 19] and gas vesicles [18, 20]), and eukaryotes (e.g. vault particles [21]). Microcompartment function depends crucially on the amount and composition of encapsulated cargo, and the integrity of the surrounding shell. Thus, understanding the mechanisms that control cargo encapsulation and shell assembly are important to elucidate the normal biological functions of microcompartments, and to reengineer microcompartments as customizable nanoreactors for synthetic biology applications (e.g. [12, 22–35]).

More broadly, there are strong parallels between the condensed cargo within a microcompartment and liquid-liquid phase-separated fluid domains. For example, the rubisco complex within a carboxysomes has a similar structure and function as the pyrenoid [36, 37]. Moreover, experiments and models suggest that microcompartment assembly proceeds by phase separation of the cargo, either prior to or during assembly of the outer shell [38–46]. However, the protein shell of a microcompartment confers distinct advantages over uncoated liquid domains. In addition to functioning as a selectively permeable barrier [47–49], experiments suggest that the shell plays a structural role, influencing the size and morphology during microcompartment assembly [23, 26, 38, 50–52]. Despite intense recent investigations into LLPS, we have yet

to understand the role of LLPS in promoting or guiding self-assembly and cargo encapsulation, or conversely the role of assembly in promoting and regulating LLPS. Thus, models that describe coupled LLPS and microcompartment assembly are needed to identify the factors that control microcompartment formation, cargo encapsulation, and function. From a practical perspective, condensation of cargo into fluid domains is a powerful approach to develop self-assembling and self-packaging gene/drug delivery vehicles [53–63]. Coupling such complexes to shell assembly could provide a means to control their size, amount of packaged cargo, and targeting capabilities.

Microcompartment shells are polydisperse (40–600 nm diameter), roughly icosahedral, and formed from pentameric, hexameric, and pseudo-hexameric (trimeric) protein oligomers [8, 9, 17, 51, 64–68]. Recently, Sutter et al. obtained atomic-resolution structures of small (40 nm diameter) empty microcompartment shells composed of pentamer, hexamer, and pseudo-hexamer proteins [51] or smaller shells (25 nm, 21 nm) composed of only pentamer and hexamer proteins [68]. While these structures provide essential insight into microcompartment architectures, the effect of cargo on shell assembly, and the mechanisms that control cargo packaging, remain unclear. In many types of microcompartments, cargo coalescence and cargo-shell interactions are mediated by auxiliary microcompartment proteins called scaffold proteins. Scaffold proteins are typically flexible (e.g. the α -carboxysome scaffold (CsoS2) is an intrinsically disordered protein) and contain multiple interaction sites with cargo molecules [37, 41, 43]. Thus, scaffolds drive cargo coalescence via weak multivalent interactions, paralleling mechanisms that drive liquid-liquid phase separation in cells.

Previous modeling studies have elucidated the assembly of empty icosahedral shells [69–95], the effect of an interior template such as a nanoparticle or RNA molecule on shell assembly [76, 82–84, 96–118], and how the interplay between

FIG. 1. Description of the computational model. (A) Bottom and (B) side views of shell subunits. Each subunit contains ‘Attractors’ (cyan spheres) on its perimeter, and a ‘Top’ (orange sphere, ‘T’) and a ‘Bottom’ (white sphere, ‘BH’) in the center above and below the hexamer plane. Interactions between Attractors drive subunit assembly, while Top-Top and Bottom-Bottom repulsions control the subunit-subunit angle and spontaneous curvature radius of the shell R_0 . (C) Scaffolds are bead-spring polymers with three domains: scaffold-cargo binding domain (yellow) beads have attractive interactions with cargo particles, middle domain (green) beads have no attractive interactions with cargo or shell subunits, and scaffold-shell binding domain (dark blue) beads have attractive interactions with Hexamer Bottom pseudoatoms ‘BH’. The contour length of the scaffold with N_s beads is denoted as L_s , while the lengths of the domains are L_{sc} , L_{sm} , and L_{sh} , respectively. Excluder atoms (orange pseudoatoms in the plane of the ‘Top’) experience excluded volume interactions with the cargo and scaffold beads. (D) Example of a complete shell with spontaneous curvature radius $R_0 = 22$, with encapsulated cargo and scaffold molecules with $L_s = 64$ ($L_{sc} = 7$, $L_{sm} = 50$, $L_{sh} = 7$). All lengths in this article are given in units of r_* , the cargo diameter (the rubisco diameter in carboxysome is ≈ 13 nm), and energies are in units of the thermal energy $k_B T$.

template curvature and shell elasticity can control its size [76, 104, 105]. However, the many-molecule scaffold-cargo complex within a microcompartment does not have a specific size or morphology, and thus requires new models. Recent computational works investigated assembly of microcompartment shell subunits [119, 120] or assembly of shells around a single cargo species with direct pairwise attractions driving cargo coalescence and packaging [44–46]. These studies showed that the cargo can influence the size of the assembled shell through a kinetic mechanism [46] or a combination of kinetic and thermodynamic effects [44, 45]. However, there are no models for microcompartment assembly that account for the crucial role of scaffold molecules.

Therefore, in this work we develop a computational model for microcompartment assembly that explicitly accounts for flexible scaffold molecules and scaffold-mediated cargo coalescence and packaging. We are motivated by recent experiments that investigated how changing the length of the

scaffold molecule affects the size of α -carboxysome shells [43, 121], but the model is sufficiently minimal and general to also provide insights into self-assembling delivery vehicles (e.g. [53–62]). We also develop an equilibrium theory, which extends models for spherical micelles assembled from star or diblock copolymers [122–126], to describe variable packaging of the interior block, and an exterior shell with a preferred curvature radius (which may differ from the preferred size of the micelle).

Our models predict that shell sizes increase with scaffold length, but saturate at minimal and maximal shell sizes in the limit of short or long scaffold molecules. Cargo packaging diminishes with increasing scaffold length (for fixed scaffold-cargo interaction valence), eventually leading to sparsely filled shells. However, increasing the scaffold-cargo valence can restore full shells.

A scaling analysis suggests that microcompartment properties are determined by the relative values of characteristic length scales set by: (i) the scaffold length, (ii) the spontaneous curvature of the protein shell, (iii) shell elasticity, and (iv) excluded volume. Our results show how these characteristic scales could be inferred from experimental measurements of the variation of shell size with scaffold length. Importantly, such measurements could reveal the role of shell spontaneous curvature in driving microcompartment closure, which is currently a key unresolved question.

I. RESULTS AND DISCUSSION

A. Computational Results

To simulate scaffold-mediated microcompartment assembly, we extend a model with direct cargo-cargo and cargo-shell pair interactions [45] to include scaffold proteins (section Computational model). Although we keep the model minimal to understand general principles, we are motivated by recent experiments on α -carboxysome assembly [43, 121]. The α -carboxysome scaffold, CsoS2, contains three linearly connected flexible domains which include binding sites for the cargo molecule rubisco and the shell hexamer proteins. While it is known that the N-terminal domain contains multiple binding sites for rubisco, as yet there is not strong experimental evidence for which sites interact with hexamers. In many other microcompartment types, the scaffold molecules contain ‘encapsulation peptides’ at their C-terminal domains that interact with hexamers, but these have not been identified for CsoS2. Given the current level of experimental uncertainty, we have started with a simple model for the scaffold: a flexible bead-spring polymer with three domains (section Computational model and Fig. 1). The domain at one end has multiple cargo interaction sites, the domain at the other end binds to shell hexamers, and the middle domain has only repulsive interactions with cargo and shell proteins. In future work we will explore other arrangements of interaction sites on the scaffold. To minimize the number of parameters, we restrict the computational model to a single size of cargo particles and scaffold segments. Thus, the simulations do not ac-

FIG. 2. Snapshots from simulation trajectories illustrating typical assembly pathways. (A) One-step assembly pathway: Coalescence of the cargo and scaffold occurs concomitantly with shell assembly. The mean size of assembled shells (defined as the radius of gyration of shell subunits) is $R = 6.3 \pm 0.5$. Parameter values are: shell spontaneous curvature radius $R_0 = 8.0$, scaffold contour length $L_s = 24$ ($L_{sm} = 10$, $L_{sc} = 7$), and hexamer-hexamer interaction strength $\epsilon_{hh} = 2.85$ (all lengths and energies are given in units of r_s and $k_B T$ respectively). (B) One-step assembly, with mean shell size $R = 18 \pm 3.6$, for $R_0 = 22$, $L_s = 64$ ($L_{sm} = 50$, $L_{sc} = 7$), and $\epsilon_{hh} = 2.65$. (C) Two-step assembly pathway: The scaffold and cargo phase separate prior to shell assembly, due to increased valence of the scaffold-cargo binding domain $f_{sc} = 0.42$. The shell size is $R = 10.66 \pm 0.60$. Parameter values are: $R_0 = 22$, $L_s = 64$ ($L_{sm} = 30$, $L_{sc} = 27$), and $\epsilon_{hh} = 2.65$. For (A)-(C), the scaffold-shell and scaffold-cargo interaction strengths are $\epsilon_{sh} = 2.5$ and $\epsilon_{sc} = 1.0$. Snapshots are shown at different scales for clarity.

count for the small excluded volume size of a CsoS2 segment relative to a rubisco holoenzyme. However, we analyze effects of varying excluded volume using the theoretical model described below.

Based on AFM experiments showing that microcompartment shell facets assemble from pre-formed hexamers [127], our model considers hexamers as the basic assembly units. Carboxysomes assembled in the presence and absence of pentamer proteins have the same structures (except that the latter presumably have vacancies at their vertices) [128]. Thus, to focus on the influence of scaffold molecules on microcompartment size, we employ a minimal model for shell assembly which includes only hexamer proteins. The hexamers have attractive edge-edge interactions with a preferred angle that sets the intrinsic spontaneous curvature radius R_0 of the shell. To further focus on scaffold-mediation, we restrict consideration to purely repulsive direct cargo-cargo and cargo-shell interactions, although attractive cargo-cargo and cargo-shell interactions have not been ruled out in the experimental systems [43, 129]. Moreover, we perform simulations with excess shell proteins to minimize effects of their depletion as

assembly proceeds. Additional model details are in the Computational model section.

Assembly pathways depend on scaffold-cargo interactions. To elucidate assembly pathways and outcomes, we performed dynamical simulations over a range of scaffold lengths L_s , cargo-binding domain lengths L_{sc} , shell spontaneous curvature radii R_0 , and binding affinities between scaffold-cargo, scaffold-hexamer, and hexamer-hexamer pairs (ϵ_{sc} , ϵ_{sh} , ϵ_{hh}). We find that the strength of scaffold-mediated cargo-cargo interactions strongly influences assembly pathways. Strong interactions ($\epsilon_{sc} \gtrsim 1.2$ for the concentrations and wild-type-length cargo binding domain length used in our simulations) drive phase separation into a dilute phase and a high density liquid complex of scaffold and cargo molecules. This condition favors two-step assembly pathways, in which the shell proteins adsorb onto the scaffold-cargo domain, and then cooperatively into ordered shell structures. For weaker scaffold-cargo interactions, we observe one-step assembly pathways with simultaneous scaffold-cargo coalescence and shell assembly (Fig. 2 A,B).

Here we are motivated by α -carboxysomes, for which ex-

FIG. 3. (A) The mean shell radius as a function of the scaffold length calculated from Brownian dynamics simulations (symbols) and the equilibrium theory (Eq. 1, lines). Results are shown for different shell spontaneous curvature radii: $R_0 = 4.5$ (●), 8.0 (▲), 22.0 (■). Example snapshots of assembled shells are shown to the left, taken from simulations with (I): $R_0 = 22$, $L_s = 14$, (II): $R_0 = 8.0$, $L_s = 24$, (III): $R_0 = 22$, $L_s = 34$, and (IV): $R_0 = 4.5$, $L_s = 84$, (V): $R_0 = 22$, $L_s = 64$. (B) Dependence of scaffold packaging on scaffold length and shell spontaneous curvature. The scaffold surface density (or ratio of packaged scaffold molecules to shell hexamer subunits, $\sigma_s \equiv n_{scaf}/n_{shell}$) is shown for the same parameter values as in (A). In (A) and (B) the length of the middle domain of the scaffolds L_{sm} is varied, with fixed shell-interacting and cargo-interacting domain lengths $L_{sh} = 7$ and $L_{sc} = 7$. The lines correspond to numerical minimization of Eq. 1, with respect to $L_s \sim R_{scaf}^2$ and σ , with the scaffold chemical potential as an adjustable parameter set to $\Delta\mu = -2.5k_B T$. Other theory parameters are taken from the simulations: $a\kappa = 12.5k_B T r_*^2$ (approximately determined in Ref. 45) and scaffold bead excluded volume $v = 0.065r_*^3$. (C) Dependence of the volume fraction of encapsulated particles (cargo and scaffolds) on the fractional length of cargo-binding domain of scaffolds $f_{sc} = L_{sc}/L_s$, for $R_0 = 22$, $L_s = 64$, and $L_{sh} = 7$. Snapshots of the interior of shells assembled at different parameter values are shown on the right: (V): $f_{sc} = 0.11$, (VI): $f_{sc} = 0.27$, (VII): $f_{sc} = 0.58$. Other parameter values for (A), (B) and (C) are as follows. Shell subunit-subunit affinities: $\epsilon_{hh} = 3.15$ at $R_0 = 4.5$, $\epsilon_{hh} = 2.85$ at $R_0 = 8.0$ and $\epsilon_{hh} = 2.65$ at $R_0 = 22$; scaffold-shell interaction $\epsilon_{sh} = 2.5$, and scaffold-cargo interaction $\epsilon_{sc} = 1.0$.

periments observe one-step pathways and that the rubisco does not undergo phase separation in the absence of shell proteins at physiological conditions [41, 67, 130]. Therefore, for all further results we set $\epsilon_{sc} = 1.0$, which drives one-step assembly pathways for the wild-type scaffold and cargo-binding domain lengths.

Examples of typical assembly trajectories are shown for scaffolds with: all domains at wild-type lengths, a long middle domain, and a long cargo binding domain in Fig. 2A,B,C respectively. For the first two examples, assembly begins with

a small aggregate involving all of the microcompartment constituents: shell subunits, scaffold, and cargo. Because we focus on weak binding affinities, all of these interactions are required to stabilize small aggregates. Consequently, the shell grows concomitant with further coalescence of scaffold and cargo, until shell closure terminates assembly. While the scaffold and cargo are essential for nucleating assembly in both cases, the scaffolds with longer middle domains (which do not attract cargo particles in the model) lead to a lower density of packaged scaffold and cargo (Fig. 2B) compared to the

shorter middle domains (Fig. 2A). In contrast, the higher valence cargo-binding domain (Fig. 2C) mediates cargo-cargo attractions that are nearly strong enough to drive phase separation, resulting in pathways closer to two-step assembly and full shells. This result is consistent with experimental observations that α -carboxysome scaffold and rubisco can phase separate without shell proteins under non-physiological conditions [37, 43].

Shell size depends on scaffold length. We find that shell size depends most sensitively on scaffold length and shell spontaneous curvature. Changing the cargo-binding domain length (at fixed total scaffold length) or binding affinities has only a small effect on shell size (Figs. S3C and S5). Motivated by the experiments, we vary the scaffold length above and below its wild-type length, with $L_s = 24$ and $L_{sc} = 7$ model segments roughly corresponding to the native values for the total scaffold length and cargo binding domain length (see SI section S2. Computational Model Details).

Fig. 3A shows the mean size of assembled shells as a function of the middle domain length L_{sm} (with two other domains at constant lengths $L_{sc} = 7$, $L_{sh} = 7$) for several values of R_0 . We identify two regimes, depending on whether the preferred size of the scaffold molecules (i.e. their end-to-end distance R_{scaf} in the absence of shell proteins, (Fig. S2)) is smaller or larger than the intrinsically preferred shell size R_0 . For small scaffold lengths, $R_{scaf} < R_0$, the shell size increases with scaffold length, consistent with experimental observations [121], until eventually saturating at a value on the order of R_0 . In this limit, the number of encapsulated scaffold molecules (n_{scaf}) relative to shell subunits (n_{shell}), given by $\sigma_s = n_{scaf}/n_{shell}$, decreases with increasing scaffold length (Fig. 3B). Finally, for $R_{scaf} \gg R_0$ shells are unable to close around scaffold molecules, leading to incomplete assembly.

Although the total amount of packaged cargo increases with scaffold length for small lengths, due to the increasing shell size, the volume fraction of packaged cargo and scaffolds decreases monotonically with scaffold length (Fig. S3A and B). I.e., long scaffold middle domains lead to poor cargo loading, because the additional scaffold segments do not attract cargo but increase excluded volume in our model.

Longer cargo binding domains increase cargo loading. Fig. 3C shows how cargo packaging depends on increasing the fractional length of the cargo binding domain, $f_{sc} = L_{sc}/L_s$; i.e., increasing L_{sc} at fixed L_s . We see that packaging increases monotonically with f_{sc} , but there is an inflection point at $f_{sc} = 0.3$, beyond which shells are nearly full. This point corresponds to the cargo binding domain length beyond which phase separation and two-step assembly pathways occur, suggesting that both thermodynamic and kinetic factors can influence cargo packaging.

Effect of binding affinity. Finally, we note that the shell size is relatively insensitive to the hexamer-hexamer ε_{hh} and scaffold-hexamer ε_{sh} affinities, provided they are strong enough to drive assembly and scaffold packaging (Fig. S5). However, the scaffold-cargo complex mediates effective hexamer-hexamer interactions, and thus larger values of ε_{hh} are required for complete shell assembly in the large L_s limit where few scaffold molecules are packaged.

B. Equilibrium theory

Model

To further understand the interplay between scaffold and shell properties, we developed a simple equilibrium model (see SI section S1. Equilibrium Theory). Although kinetic factors may be important as noted above, the thermodynamic model captures key trends observed in the dynamical simulations.

We make the following simplifying assumptions; extensions to eliminate these approximations are straightforward. First, we restrict the theory to spherical shells as observed in the simulations. Second, cargo molecules are implicitly represented by a net driving force for scaffold packaging $\Delta\mu$ that accounts for scaffold-shell and scaffold-cargo attractive interactions, as well as the mixing entropy penalty for scaffold coalescence. We further assume that the coalesced cargo and scaffold-cargo binding domains are concentrated in the shell interior, while the other end of the scaffold, the scaffold-shell binding domain, binds to the inner shell surface. Our model system can thus be thought of as a diblock copolymer spherical micelle (e.g. [122–124, 131–133]), except that the stoichiometry of the interior/exterior block is annealed, and there is an exterior shell with a preferred curvature that may be incommensurate with the optimal size of the micelle within. We discuss below the implications of model predictions for alternative scaffold packaging geometries.

The free energy (per shell subunit) of a shell with radius R and ratio of scaffold/shell subunits σ_s is given by

$$f_{\text{assem}} = f_{\text{shell}} + \sigma_s f_{\text{scaf}} + f_{\text{ent}} \quad (1)$$

$$f_{\text{shell}} = g_{hh} + \frac{a\kappa}{2} \left(\frac{2}{R} - \frac{2}{R_0} \right)^2 \quad (2)$$

$$f_{\text{scaf}} = \left(\frac{R}{R_{\text{scaf}}} \right)^2 + \left(\frac{R_{\text{scaf}}}{R} \right)^2 + \Delta\mu + \frac{3}{2} \frac{\nu \sigma_s N_s^2}{aR} \quad (3)$$

$$f_{\text{ent}} = \sigma_s \log \sigma_s + (1 - \sigma_s) \log(1 - \sigma_s) \quad (4)$$

Here f_{shell} gives the shell free energy, with the first term (g_{hh}) the free energy per hexamer due to hexamer-hexamer contacts (see SI section S1. Equilibrium Theory, Eq. S8), and the second term describing the bending energy due to deviations from the shell spontaneous curvature, with a the subunit area and κ the bending modulus. For simplicity, we neglect contributions from the 12 five-fold defects that are required by topology (in physical and computational microcompartments). Including these will not qualitatively change the results (see SI section S1. Equilibrium Theory and Ref. 45). Then the first two terms in f_{scaf} give the entropic penalty for stretching or compressing scaffold molecules from R_{scaf} . The following two terms account for attractive scaffold-cargo and scaffold-shell interactions ($\Delta\mu$), and excluded volume interactions, with ν as the scaffold segment excluded volume. Finally, f_{ent} represents the mixing entropy of encapsulated scaffolds. The equilibrium shell size and scaffold loading are obtained by minimizing the total free energy of the system. In the thermodynamic limit, the distribution of shell sizes will be

strongly peaked around the size that minimizes the free energy **per subunit** of the shell complex (see SI section S1. Equilibrium Theory and Refs. [45, 134]).

Comparison of theory with computational results

The solid lines in Fig. 3A,B show the theory results, with the parameters $a\kappa = 12.5 k_B T r_*^2$, and $\nu = 0.065 r_*^3$ calculated from simulations and $\Delta\mu$ as a fitting parameter chosen by eye. We set $R_{\text{scaf}} = l_k N_s^{1/2}$ since the theory separately accounts for chain stretching and excluded volume (as in the Flory approximation [135]).

We find that the theory qualitatively captures the main features observed in the simulation results. In particular, the shell size increases with scaffold length, until saturating near R_0 for long scaffolds, where excluded volume limits scaffold packaging. However, for simulations with $R_0=22$, longer scaffolds undergo limited coalescence in the center of the shell, causing the shell size to saturate at smaller scaffold lengths than predicted by the theory.

Scaling analysis, and the small excluded volume limit

Eq. 1 predicts that the system behavior depends on several characteristic length scales. In addition to the shell spontaneous curvature radius R_0 and scaffold size R_{scaf} , we obtain length scales from the scaffold excluded volume, $R_{\text{exc}} \sim (N_s \nu)^{1/3}$, and the shell elastic energy, $R_E = \sqrt{2a\kappa}$. Specifically, R_E is the radius for which a shell with no spontaneous curvature $R_0 = \infty$ would have a bending energy of $k_B T$ per subunit. Different scaling forms can be deduced from Eq. 1 depending on the relative magnitudes of these lengths.

Since our model is most applicable to biological microcompartments in the limit of small excluded volume (see discussion), we focus here on the limit of small excluded volume, $\nu \rightarrow 0$. In SI section S1. Equilibrium Theory we present a detailed derivation of these expressions and further analysis for finite excluded volume.

Fig. 4 shows the shell size calculated as a function of scaffold length and shell spontaneous curvature with $\nu = 0$, by minimizing Eq. 1. The dashed black lines encompass the region with significant scaffold packaging for $\Delta\mu = -10k_B T$; the full dependence of scaffold packaging on parameters is shown in Fig. S1. We see that the range of accessible shell sizes depends crucially on the relative sizes of the preferred shell curvature and the elastic length scale. For $R_0 > R_E$ the shell size is highly tunable with scaffold length, varying over orders of magnitude, while only small deviations from the preferred shell curvature are possible in the opposite limit $R_0 < R_E$. This predicted trend may be a way to experimentally probe R_0 in microcompartment systems.

Scaling analysis. Further insight can be obtained from a simple scaling analysis. First, minimizing the free energy with respect to σ_s gives

$$\sigma_{\text{eq}} \approx 1 / \left(1 + e^{f_{\text{scaf}}(R_{\text{eq}})} \right). \quad (5)$$

Eq. 5 shows that scaffold/cargo packaging diminishes as the shell size deviates from the preferred scaffold size R_{scaf} , but that scaffold can be packaged over a greater range of shell sizes by increasing $|\Delta\mu|$. Second, we obtain simplified expressions for the shell size and packaging depending on the relative characteristic scales as follows.

Small intrinsic shell curvature, $R_0 \gg R_E$. In this limit the optimal shell size becomes asymptotically independent of shell spontaneous curvature, and minimizing the free energy results in an optimal shell size

$$R_{\text{eq}} \approx R_{\text{scaf}} \left[1 + \frac{1}{\sigma_{\text{eq}}} \left(\frac{R_E}{R_{\text{scaf}}} \right)^2 \right]^{1/4} \quad \text{for } R_0 \gg R_E, \quad (6)$$

which can be further simplified in two limits:

$$R_{\text{eq}} \approx \begin{cases} R_{\text{scaf}} \left[1 + \frac{1}{\sigma_{\text{eq}}} \left(\frac{R_E}{2R_{\text{scaf}}} \right)^2 \right] \approx R_{\text{scaf}} & \text{for } R_{\text{scaf}} \gg R_E \\ (R_{\text{scaf}} R_E)^{1/2} \sigma_{\text{eq}}^{-1/4} & \text{for } R_{\text{scaf}} \ll R_E \end{cases} \quad (7)$$

Thus, the shell size depends on a competition between the preferred scaffold size and the tendency of the shell proteins to disfavor curvature: the shell size tracks the scaffold size $R_{\text{eq}} \approx R_{\text{scaf}}$ for long scaffolds, but adopts the geometric mean of the scaffold size and shell elastic length for short scaffolds. These scaling estimates are compared against the full free energy in Fig. 5A.

However, Eq. 5 shows that scaffold packaging will exponentially diminish for large deviations from R_{scaf} . Therefore, we estimate upper and lower bounds on the scaffold-driven changes in shell size by noting that shell bending energy must be compensated by the free energy of scaffold packaging:

$$R_{\text{eq}}^{\text{max}} \approx \begin{cases} \frac{R_0 R_E}{R_E - \Delta\mu_{\text{eff}}^{1/2} R_0}, & \text{for } \Delta\mu_{\text{eff}}^{1/2} < \frac{R_E}{R_0} \\ \infty, & \text{for } \Delta\mu_{\text{eff}}^{1/2} \geq \frac{R_E}{R_0} \end{cases} \quad (8)$$

$$R_{\text{eq}}^{\text{min}} \approx \frac{R_0 R_E}{R_E + \Delta\mu_{\text{eff}}^{1/2} R_0} \approx R_E / \Delta\mu_{\text{eff}}^{1/2}, \quad \text{for } R_0 \rightarrow \infty \quad (9)$$

with $\Delta\mu_{\text{eff}} \equiv -(\Delta\mu + 2k_B T)\sigma_{\text{eq}} - f_{\text{ent}}$ the effective confinement free energy per shell subunit, since $f_{\text{scaf}} \approx \Delta\mu + 2k_B T$ for $R \approx R_{\text{scaf}}$, and

$$\sigma_{\text{eq}} \approx 1 / \left(1 + e^{\Delta\mu + 2k_B T} \right). \quad (10)$$

The packageable range of scaffold lengths can then be determined by combining Eqs. 8, 9 and 7 (Fig. 4B)). Outside of these bounds, scaffold packaging decreases exponentially, resulting in either no assembly or assembly of empty shells with sizes near R_0 . Alternatively, for small scaffolds, if the scaffold-shell energy is sufficient to drive scaffold packaging without cargo, the scaffold can form a layer in the vicinity of the shell surface (Fig. 4C). That scenario is not accounted for in the theoretical model.

FIG. 4. (A) Shell radius as a function of the shell spontaneous curvature radius R_0 and scaffold unperturbed size R_{scaf} (normalized by the shell elastic energy length scale R_E), calculated by numerically minimizing Eq. 1, with scaffold chemical potential $\Delta\mu = -10k_B T$, $R_E = 30r_*$, subunit area $a = 2.5r_*^2$, and excluded volume parameter $\nu = 0$. The dashed line encompasses parameter values that lead to efficient scaffold packaging (see Fig. S1 for the amount of packaged scaffold as a function of these parameters). Outside of this region $\sigma_s \approx 0$, leading to assembly of nearly empty shells (with a surface layer of scaffold and cargo) with sizes approximately equal to R_0 , incomplete shells, or no assembly. The heat map in this region corresponds to the size of the complete but nearly empty shells. (B) The minimum and maximum shell radius (normalized by R_E) predicted by the theory for varying the scaffold length as a function of the scaffold packaging driving force $\Delta\mu$, for fixed shell spontaneous curvature $R_0 = 50$. The left y-axis shows the shell size, while the right y-axis shows the corresponding scaffold length, showing that the shell size closely tracks the scaffold preferred size in this regime. The symbols correspond to maximum and minimum shell sizes obtained by numerically minimizing Eq. 1, and the lines show the asymptotic results (Eq. 7). For scaffold lengths above the maximum length, shells would be either incomplete or empty, whereas below the minimum length, shells will either be incomplete or have low packaging of scaffold and cargo. Other parameters are the same as in (A). (C) Examples of incomplete assemblies that form in the dynamics simulations with overly short (top) or long (bottom) scaffold molecules. Parameters are (top) $R_0 > 300$ and $L_s = 34$, (bottom) $R_0 = 4.5$ and $L_s = 84$. Other parameters are as in Fig. 3.

Importantly, in any of these scenarios, the microcompartments will saturate at a *minimum* shell size with decreasing scaffold length (Eq. 9), which is an experimentally testable prediction.

High intrinsic shell curvature, $R_0 < R_E$. In this regime the preferred curvature of the shell proteins influences the assembled shell size, so we must consider the full free energy Eq. 1. Then, the limits in which the preferred scaffold size is larger or smaller than R_0 result in

$$R_{eq} \approx \begin{cases} R_0 \left[1 + \sigma_{eq} (R_{scaf}/R_E)^2 \right], & \text{for } R_{scaf} \gg R_0 \\ R_0 \left[1 - \frac{\sigma_{eq} R_0^4}{R_{scaf}^2 R_E^2} \right], & \text{for } R_{scaf} \ll R_0 \end{cases} \quad (11)$$

showing that the shell size depends on the competition between the shell spontaneous curvature and scaffold size, with relative importance determined by the shell bending modulus (via R_E). As above, there are bounds on the range of scaffold lengths that can be efficiently packaged, which increase with $|\Delta\mu|$. Since the shell bending energy dominates in this limit, the shell size exhibits only small deviations from R_0 , and the extent of scaffold incorporation is determined by the competition between the packaging driving force $\Delta\mu$ and the scaffold confinement free energy f_{scaf} , resulting in:

$$R_{scaf}^{max} \approx R_0 \left[\frac{1}{\Delta\mu_{eff} - 1} - 2\sigma_{eq} \left(\frac{R_0}{R_E} \right)^2 \right]^{-1/2}$$

$$R_{scaf}^{min} \approx R_0 \left[\frac{1}{\Delta\mu_{eff} - 1} + 2\sigma_{eq} \left(\frac{R_0}{R_E} \right)^2 \right]^{1/2}. \quad (12)$$

The range of shell sizes ($R_{eq}^{min}, R_{eq}^{max}$) can then be determined by combining Eqs. 10-12. Consistent with the numerical solution (Figs. 4 and S1), the range of accessible shell sizes is considerably smaller than in the large R_0 limit.

II. CONCLUSIONS

We have theoretically and computationally investigated the equilibrium behavior and dynamical assembly pathways of a system in which polymeric scaffold molecules drive coalescence of many cargo particles, and their packaging within an assembling protein shell. While motivated by biological systems such as bacterial microcompartments, we have employed minimal models allowing for generic conclusions about such multicomponent assemblies. Thus, in addition to elucidating the behaviors of natural and reengineered bacterial microcompartments, our results suggest design principles for tailoring the size, morphology, and loading of synthetic or biomimetic systems such as delivery vehicles.

We find that the shell size and the amount of packaged cargo are determined by several characteristic length scales: the preferred sizes of the protein shell and the scaffold, the scaffold excluded volume size, and a characteristic elastic length scale of the shell (which depends on its rigidity). In our computational model, we have considered a relatively low bending rigidity of the shell (motivated by experimental observations on bacterial microcompartments), and scaffold excluded volume plays a significant role. In this case, results depend on whether the preferred scaffold size is ei-

FIG. 5. Dependence of shell size on the scaffold preferred size R_{scaf} calculated from the asymptotic analysis, (lines, Eqs. (7) and (11), compared against numerical minimization of Eq. 1 (symbols). Results are shown in (A) and (B) respectively for shell spontaneous curvature radius values that are large or small compared to the shell elastic length scale: (A) $R_0 = 300$ and (B) $R_0 = 1$, with $R_E = 30$. Other parameter values are $a = 2.5r_*^2$, $\Delta\mu = -10k_B T$, and $\nu = 0$.

ther smaller than or larger than the spontaneous curvature radius. In the limit of small scaffold molecules, assembly driving forces are dominated by stretching of the scaffold between the cargo-filled interior and the surface of the protein shell, leading to shells which are smaller than the spontaneous curvature radius (provided that the scaffold-shell interactions and scaffold-mediated cargo-cargo interactions are stronger than shell bending elasticity). In the opposite limit of long scaffold molecules, scaffold packing becomes a dominant contribution, leading to decreased scaffold-cargo packaging and a shell size that roughly saturates at the spontaneous curvature radius.

Our simulations also demonstrate that the amount and morphology of packaged cargo depends on the strength and geometry of scaffold-cargo interactions. Multivalent scaffold-cargo interactions are essential for achieving full shells, while the scaffold length and distribution of cargo binding sites on the scaffold influence the morphology of the packaged cargo. When cargo-binding sites are limited to one end of long scaffold molecules, with scaffold-shell interactions at the other end, we observe a tri-layer micelle-like complex, with an interior core of cargo, surrounded by a layer of scaffold, and finally the shell on the exterior. In contrast, longer scaffold-cargo binding domains or shorter scaffolds lead to a nearly uniform distribution of cargo and scaffold in the shell interior.

Because our primary objective was to qualitatively understand scaffold-mediated shell assembly, we did not tune the computational model parameters to match a specific biological system, and the computational model likely overestimates the scaffold excluded volume compared to bacterial microcompartments. For example, the α -carboxysome scaffold (CsoS2) accounts for only 12% by mass of a native α -carboxysome [136], compared to 70% in our model. Moreover, while we have included very weak scaffold inter-segment attractions in our model, stronger inter-segment attractions among CsoS2 residues could further reduce the effective excluded volume.

To facilitate more direct comparisons against experiments, we also developed a more tractable, though simplified, equilibrium theoretical model. Using this model, we have explored a wide parameter range, including the effect of reducing the excluded volume. The theoretical analysis suggests that the ability of a scaffold to modulate shell size depends crucially on whether the protein shell spontaneous curvature radius is large or small compared to its elastic length scale. In the regime of a large spontaneous curvature radius, shell size is extremely malleable and can be tuned over a wide range by varying scaffold length. In contrast, shell proteins with a small preferred curvature radius exhibit only small size fluctuations, and relatively poor packaging of scaffold molecules with incommensurate sizes. In both regimes, the range of scaffold lengths that can be packaged effectively, and correspondingly the range of accessible shell radii, increase with the free energy driving force for scaffold incorporation. This important control parameter depends on the scaffold concentration and the valence and affinity of the scaffold-cargo and scaffold-shell domains.

Testing against experiments. Since our models are general, these predictions could be tested in a variety of biological or synthetic systems. However, bacterial microcompartments are the existing system in which testing would be most straightforward, especially due to the modular nature of microcompartment scaffold molecules. For example, the CsoS2 scaffold of the α -carboxysome can be represented as a tri-block co-polymer, with the cargo-binding domain in the N-terminal block, and shell-binding sites in the C-terminal block. Our results could thus be compared against experiments in which mutagenesis [43] is used to change the length of the middle domain (Fig. 3A,B) or the cargo-binding domain (Fig. 3C), although we note that the presence of shell binding sites in the middle block has not been ruled out.

Importantly, our results suggest a path to assess the magnitude and importance of the spontaneous curvature, a key unresolved question in microcompartment assembly. In particular, recent experiments on recombinant systems exhibit assembly of small ($\sim 20-30$ nm) empty microcompartment shells, suggesting a small shell spontaneous curvature radius. Yet, other experiments observe extensive flat sheets of hexamers assembling on surfaces, suggesting at least the possibility of a low or zero spontaneous curvature. In this regard, both the simulations and theoretical model predict that shells with a small spontaneous curvature radius exhibit only narrow fluctuations in size when presented with scaffold in molecules of differ-

ent lengths, whereas shells with a large spontaneous curvature radius can assemble over a wide range of sizes to accommodate scaffolds with different lengths. In the small spontaneous curvature regime, our results suggest that preferred shell curvature plays a key role in driving shell closure, whereas in the large-curvature radius regime scaffold properties (this work) and cargo properties [45, 46] are important for shell closure.

One readily testable prediction from our model is that there is a minimum accessible shell size. Reducing scaffold lengths below this limit will either lead to a saturating shell size with diminished scaffolding packaging, or abrogation of shell assembly.

Outlook. Finally, our results suggest additional physical ingredients that will be important to study for understanding biological microcompartments or guiding the design of synthetic delivery vehicles. First, increasing the rigidity of the scaffold molecules would diminish the range of variation of shell sizes. This trend can be seen from the equilibrium scaling of the shell size in Eqs. 7 and 11, since the unperturbed scaffold size increases with its rigidity. Second, combining direct cargo-cargo interactions with scaffold mediation could enhance cargo packaging within assembling shells, since. This proposition is based on our observation in this work that increasing effective scaffold-mediated cargo interactions promotes cargo loading, and is consistent with observations from previous models with cargo coalescence driven only by direct cargo-cargo pair interactions [44]. Third, although we observe qualitative agreement between our dynamical simulation results and the equilibrium theory, kinetic effects likely play important roles in some parameter regimes. For example, we have primarily focused on parameters that lead to one-step assembly pathways as observed for α -carboxysomes, but we anticipate that kinetic effects will become more important in the case of two-step assembly pathways. Thus, models such as developed here, combined with experiments on different microcompartment systems and solution conditions, are needed to fully understand the interplay between thermodynamics and kinetics in scaffold-mediated shell assembly and cargo packaging.

III. METHODS

A. Computational model

Our computational model extends a previous model for assembly of a shell around fluid cargo [44, 45] to include scaffold molecules and scaffold-mediated cargo-cargo and cargo-shell interactions. For simplicity, we do not include direct cargo-cargo and cargo-shell attractive interactions, and we consider only hexamer shell proteins. The interactions among components are as follows, with further details given in SI section S2. Computational Model Details.

Scaffolds. In carboxysome systems, attractions between rubisco particles are mediated by auxiliary proteins (*e.g.* the intrinsically disordered protein CsoS2 in α -carboxysomes [41, 43] and the protein CcmM in β -carboxysomes [38]). The α -carboxysome scaffold molecule, CsoS2, contains three

linearly connected flexible domains: the C-terminal domain contains multiple binding sites for the cargo molecule rubisco, while there is not yet clear experimental evidence for the location of the shell binding sequences. However, in β -carboxysomes and other bacterial microcompartments, shell-cargo attractions are mediated by an ‘encapsulation peptide’ sequence in the N-terminal domain of the scaffold [23, 31, 38, 137, 138]. Motivated by these characteristics, we model the scaffold as a flexible bead-spring polymer with three linearly connected domains: a scaffold-cargo binding domain at one end with multiple cargo interaction sites, a scaffold-shell binding domain at the other end with multiple shell hexamer binding sites, and a middle domain that has only repulsive interactions with cargo and shell proteins. Based on experimental evidence, we include weak attractive interactions between pairs of scaffold beads.

The interactions between various molecule types are modeled as follows.

Scaffold-cargo interactions. Attractive interactions between cargo particles and the ‘cargo-interacting’ beads (type ‘SC’) in scaffolds are modeled by a Lennard-Jones potential with well-depth parameter ϵ_{sc} .

Scaffold-shell interactions. Attractive interactions between hexamers and scaffolds are modeled by a Morse potential with well-depth parameter ϵ_{sc} , between beads in ‘shell interacting’ beads in polymers (blue beads in Fig. 1C) and Bottom pseudoatoms on hexamers (type ‘BH’). We also add a layer of ‘Excluders’ in the plane of the ‘Top’ pseudoatoms, which account for shell-scaffold and shell-cargo excluded volume interactions.

Shell-shell interactions. Interactions between edges of BMC shell proteins are primarily driven by shape complementarity and hydrophobic interactions [51]. To mimic these short-ranged directionally specific interactions, each model subunit contains ‘Attractors’ on its perimeter that mediate shell-shell attractions. Complementary Attractors on nearby subunits have short-range interactions (modeled by a Morse potential, Eq. S26 in SI section S2. Computational Model Details). Attractors which are not complementary do not interact. The arrangement of Attractors on subunit edges is shown in Fig. 1, with pairs of complementary Attractors indicated by cyan double-headed arrows. The shell-shell binding affinity is proportional to the well-depth of the Morse potential between complementary Attractors, given by ϵ_{hh} .

To control the shell spontaneous curvature and bending modulus, each subunit contains a ‘Top’ (type ‘TH’) pseudoatom above the plane of Attractors, and a ‘Bottom’ pseudoatom (type ‘BH’) below the Attractor plane. There are repulsive interactions (WCA interactions, Eq. S25) between Top-Top, Bottom-Bottom, and Top-Bottom pairs of pseudoatoms on nearby subunits. The relative sizes of the Top and Bottom pseudoatoms set the preferred subunit-subunit binding angle (and thus the spontaneous curvature R_0), while the interaction strength (controlled by the well-depth parameter ϵ_{angle}) controls the shell bending modulus κ . The Top-Bottom interaction ensures that subunits do not bind in inverted orientations [77]. Since the shell-shell interaction geometries are already controlled by the Attractor, Top, and Bottom pseu-

doatoms, we do not consider Excluder-Excluder interactions.

B. Simulations

We simulated assembly dynamics using the Langevin dynamics algorithm in HOOMD (which uses GPUs to efficiently simulate dynamics [139]), and periodic boundary conditions to represent a bulk system. The subunits were modeled as rigid bodies [140]. Each simulation was performed in the NVT ensemble, using a set of fundamental units [141] with the unit length scale r_* defined as the cargo particle diameter, and energies in units of the thermal energy, $k_B T$. The simulation time step was 0.005 in dimensionless time units, and we performed 2×10^8 timesteps in each simulation, except the maximum simulation time was increased to 4×10^8 for $L_s \geq 40$ and $R_0 = 22$ (because longer timescales were required for the large shells that assemble at those parameters).

Systems. Each system included 2300 hexamers, 2000 cargo particles, and 213 scaffold molecules, initialized randomly on a grid in a cubic box with side length $120r_*$. For systems that are near the threshold for cargo-scaffold phase separation in the absence of shell subunits because of long scaffold-cargo binding domains ($L_{sc} > 7$), we eliminated correlations imposed by the grid by performing an initial dynamics with excluded volume interactions only (all attractive interactions turned off) for 5×10^6 timesteps. The attractive interactions were turned on after completion of this initialization dynamics.

Sample sizes. We performed a minimum of 10 independent trials at each parameter set. Simulations with a large spontaneous curvature radius and long scaffolds, $R_0 = 22$ and $L_s > 40$, result in relatively few complete shells per simulation. Therefore, for these parameters we performed additional trials such that at least 10 complete shells formed amongst all trials at a given parameter set.

Connecting simulation nondimensional units to physical values. Although the model is designed to be generic, we are particularly motivated by α -carboxysomes. We can *approximately* map our computational model to carboxysomes by setting the cargo diameter (the unit length scale in the model) to that of the rubisco holoenzyme, implying $r_* \approx 13$ nm. However, to enable tractable simulation of long assembly timescales, we have set the size ratios of hexamer/cargo

and scaffold bead/cargo to be larger than the ratios of these proteins relative to rubisco. In particular, our model hexamers have a side length of r_* and are thus about three times larger than carboxysome hexamers (side-length ≈ 4 nm). To represent the carboxysome system as closely as possible despite this approximation, we have set the bending modulus of the computational model to obtain a characteristic elastic length scale $R_E = \sqrt{2ak}$ similar to that of carboxysomes. Nanoindentation measurements on β -carboxysome shells obtained $\kappa \approx 25k_B T$ [142] (smaller than that for a typical viral capsid), resulting in an elastic length scale of $R_E \approx 47$ nm. Correspondingly, we set the computational bending modulus to about $5 - 10k_B T$ (see SI section S2. Computational Model Details), and thus $R_E \approx 38 - 54$ nm (using a subunit area $a \approx 2.5r_*^2$).

We perform a similar approximate mapping for the model scaffold molecule, by attempting to match its end-to-end vector (when unconfined) to that of a CsoS2 molecule. The model scaffold behaves as a self-avoiding polymer with a statistical segment length of about one segment, or $l_k \approx 0.5r_*$, and thus has a end-to-end vector $R_{scaf} \approx l_k N_s^{0.58} r_*$, with $N_s = L_s/l_k$ (see Fig. S2), with a proportional radius of gyration. Since most of the CsoS2 sequence is thought to be intrinsically disordered, we assume that it also behaves as a self-avoiding polymer. The *Halothiobacillus neapolitanus* CsoS2 molecule studied in mutagenesis experiments has 869 amino acids. While its radius of gyration has not been reported, the *Prochlorococcus* CsoS2 with a similar sequence length (792 amino acids) has a radius of gyration of ~ 70 nm [41]. To match this value, we assign the wild-type length of the model scaffold to be $L_s = 25r_*$ (50 beads).

ACKNOWLEDGMENTS

We are grateful to Luke Oltrogge and David Savage for insightful discussions. This work was supported by Award Number R01GM108021 from the National Institute Of General Medical Sciences and the Brandeis Center for Bioinspired Soft Materials, an NSF MRSEC, DMR-1420382 and DMR-2011846. Computational resources were provided by NSF XSEDE computing resources (XStream, Bridges, and Comet) and the Brandeis HPCC which is partially supported by DMR-1420382.

- [1] C. P. Brangwynne, J. Cell Biol. **203**, 875 (2013).
- [2] S. C. Weber and C. P. Brangwynne, Cell **149**, 1188 (2012).
- [3] A. A. Hyman and C. P. Brangwynne, Dev. Cell **21**, 14 (2011).
- [4] A. A. Hyman and K. Simons, Science **337**, 1047 (2012).
- [5] T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, and J. M. Shively, Nat. Rev. Microbiol. **6**, 681 (2008).
- [6] T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, and J. D. Forman-Kay, Molecular cell **57**, 936 (2015).
- [7] B. Y. Zaslavsky, L. A. Ferreira, A. L. Darling, and V. N. Uversky, Int. J. Biol. Macromol. **117**, 1224 (2018).
- [8] C. A. Kerfeld, S. Heinhorst, and G. C. Cannon, Microbiology **64**, 391 (2010).
- [9] S. Tanaka, C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, and T. O. Yeates, Science **319**, 1083 (2008), <http://www.sciencemag.org/content/319/5866/1083.full.pdf>.
- [10] B. D. Rae, B. M. Long, M. R. Badger, and G. D. Price, Microbiol. Mol. Biol. Rev. **77**, 357 (2013), <http://mmbr.asm.org/content/77/3/357.full.pdf+html>.
- [11] T. A. Bobik, B. P. Lehman, and T. O. Yeates, Mol. Microbiol. **98**, 193 (2015).

- [12] C. Chowdhury, S. Sinha, S. Chun, T. O. Yeates, and T. A. Bobik, *Microbiol. Mol. Biol. Rev.* **78**, 438 (2014).
- [13] C. A. Kerfeld and M. R. Melnicki, *Curr. Opin. Plant Biol.* **31**, 66 (2016).
- [14] J. K. Polka, S. G. Hays, and P. A. Silver, *Cold Spring Harb Perspect Biol* **8** (2016), 10.1101/cshperspect.a024018.
- [15] C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, and M. Sutter, *Nat. Rev. Microbiol.* **16**, 277 (2018).
- [16] M. F. Schmid, A. M. Paredes, H. A. Khant, F. Soyer, H. C. Aldrich, W. Chiu, and J. M. Shively, *J. Mol. Biol.* **364**, 526 (2006).
- [17] C. V. Iancu, H. J. Ding, D. M. Morris, D. P. Dias, A. D. Gonzales, A. Martino, and G. J. Jensen, *J. Mol. Biol.* **372**, 764 (2007).
- [18] M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, and N. Ban, *Nat. Struct. Mol. Biol.* **15**, 939 (2008).
- [19] R. J. Nichols, B. LaFrance, N. R. Phillips, L. M. Oltrogge, L. E. Valentin-Alvarado, A. J. Bischoff, E. Nogales, and D. Savage, *bioRxiv* (2020).
- [20] F. Pfeifer, *Nat Rev Micro* **10**, 705 (2012).
- [21] V. A. Kickhoefer, K. S. Rajavel, G. L. Schaffer, W. S. Dalton, R. J. Scheper, and L. H. Rome, *J. Biol. Chem.* **273**, 8971 (1998), <http://www.jbc.org/content/273/15/8971.full.pdf+html>.
- [22] A. Chessher, R. Breitling, and E. Takano, *ACS Biomaterials Science & Engineering* **1**, 345 (2015).
- [23] B. P. Lehman, C. Chowdhury, and T. A. Bobik, *J. Bacteriol.* **199** (2017), 10.1128/JB.00785-16.
- [24] B. M. Long, M. R. Badger, S. M. Whitney, and G. D. Price, *J. Biol. Chem.* **282**, 29323 (2007).
- [25] B. M. Long, L. Tucker, M. R. Badger, and G. D. Price, *Plant Physiol.* **153**, 285 (2010).
- [26] M. J. Mayer, R. Juodeikis, I. R. Brown, S. Frank, D. J. Palmer, E. Deery, D. M. Beal, W. F. Xue, and M. J. Warren, *Sci. Rep.* **6**, 36899 (2016).
- [27] M. B. Quin, S. A. Perdue, S. Y. Hsu, and C. Schmidt-Dannert, *Appl. Microbiol. Biotechnol.* **100**, 9187 (2016).
- [28] M. Slininger Lee and D. Tullman-Ercek, *Current Opinion in Systems Biology* (2017), 10.1016/j.coisb.2017.05.017.
- [29] M. F. Slininger Lee, C. M. Jakobson, and D. Tullman-Ercek, *ACS Synth. Biol.* (2017), 10.1021/acssynbio.7b00042.
- [30] A. Hagen, M. Sutter, N. Sloan, and C. A. Kerfeld, *Nat. Comm.* **9**, 1 (2018).
- [31] J. N. Kinney, A. Salmeen, F. Cai, and C. a. Kerfeld, *J. Biol. Chem.* **287**, 17729 (2012).
- [32] B. Ferlez, M. Sutter, and C. A. Kerfeld, *Metab. Eng.* **54**, 286 (2019).
- [33] G. Kalnins, E.-E. Cesle, J. Jansons, J. Liepins, A. Filimonenko, and K. Tars, *Nat. Comm.* **11**, 1 (2020).
- [34] R. Juodeikis, M. J. Lee, M. Mayer, J. Mantell, I. R. Brown, P. Verkade, D. N. Woolfson, M. B. Prentice, S. Frank, and M. J. Warren, *MicrobiologyOpen* **9**, e1010 (2020).
- [35] M. Künzle, J. Mangler, M. Lach, and T. Beck, *Nanoscale* **10**, 22917 (2018).
- [36] J.-D. Rochaix, *Cell* **171**, 28 (2017).
- [37] H. Wang, X. Yan, H. Aigner, A. Bracher, N. D. Nguyen, W. Y. Hee, B. Long, G. D. Price, F. Hartl, and M. Hayer-Hartl, *Nature* **566**, 131 (2019).
- [38] J. C. Cameron, S. C. Wilson, S. L. Bernstein, and C. A. Kerfeld, *Cell* **155**, 1131 (2013).
- [39] A. H. Chen, A. Robinson-Mosher, D. F. Savage, P. A. Silver, and J. K. Polka, *PLoS One* **8**, e76127 (2013).
- [40] T. Wunder, S. L. H. Cheng, S.-K. Lai, H.-Y. Li, and O. Mueller-Cajar, *Nat. Comm.* **9**, 1 (2018).
- [41] F. Cai, Z. Dou, S. L. Bernstein, R. Leverenz, E. B. Williams, S. Heinhorst, J. Shively, G. C. Cannon, and C. A. Kerfeld, *Life* **5**, 1141 (2015).
- [42] T. Chaijarasphong, R. J. Nichols, K. E. Kortright, C. F. Nixon, P. K. Teng, L. M. Oltrogge, and D. F. Savage, *J. Mol. Biol.* **428**, 153 (2016).
- [43] L. M. Oltrogge, T. Chaijarasphong, A. W. Chen, E. R. Bolin, S. Marqusee, and D. F. Savage, *Nature Structural & Molecular Biology*, 1 (2020).
- [44] J. D. Perlmutter, F. Mohajerani, and M. F. Hagan, *eLife* **5**, e14078 (2016).
- [45] F. Mohajerani and M. F. Hagan, *PLoS Comput. Biol.* **14**, e1006351 (2018).
- [46] G. M. Rotskoff and P. L. Geissler, *Proceedings of the National Academy of Sciences* **115**, 6341 (2018).
- [47] B. B. Menon, S. Heinhorst, J. M. Shively, and G. C. Cannon, *Journal of bacteriology* **192**, 5881 (2010).
- [48] P. Mahinthichaichan, D. M. Morris, Y. Wang, G. J. Jensen, and E. Tajkhorshid, *J. Phys. Chem. B* **122**, 9110 (2018).
- [49] M. Faulkner, I. Szabo, S. L. Weetman, F. Sicard, R. G. Huber, P. J. Bond, E. Rosta, and L.-N. Liu, *bioRxiv* (2020).
- [50] J. K. Lassila, S. L. Bernstein, J. N. Kinney, S. D. Axen, and C. A. Kerfeld, *J. Mol. Biol.* **426**, 2217 (2014).
- [51] M. Sutter, B. Greber, C. Aussignargues, and C. A. Kerfeld, *Science* **356**, 1293 (2017).
- [52] F. Cai, S. L. Bernstein, S. C. Wilson, and C. A. Kerfeld, *Plant Physiol.* **170**, 1868 (2016).
- [53] S. Y. Fung, H. Yang, P. T. Bhola, P. Sadatmousavi, E. Muzar, M. Liu, and P. Chen, *Adv. Funct. Mater.* **19**, 74 (2009).
- [54] H. Vilaça, T. Castro, F. M. Costa, M. Melle-Franco, L. Hilliou, I. W. Hamley, E. M. Castanheira, J. A. Martins, and P. M. Ferreira, *J. Mater. Chem. B* **5**, 8607 (2017).
- [55] N. A. N. Hanafy, M. El-Kemary, and S. Leporatti, *Cancers* **10**, 238 (2018).
- [56] C. Gao, H. Li, Y. Li, S. Kewalramani, L. C. Palmer, V. P. Dravid, S. I. Stupp, M. Olvera de la Cruz, and M. J. Bedzyk, *J. Phys. Chem. B* **121**, 1623 (2017).
- [57] K. Hpone Myint, J. R. Brown, A. R. Shim, B. E. Wyslouzil, and L. M. Hall, *J. Phys. Chem. B* **120**, 11582 (2016).
- [58] M. Kang, H. Cui, and S. M. Loverde, *Soft Matter* **13**, 7721 (2017).
- [59] M. T. Lee, A. Vishnyakov, and A. V. Neimark, *J. Phys. Chem. B* **117**, 10304 (2013).
- [60] S. M. Loverde, M. L. Klein, and D. E. Discher, *Adv. Mater.* **24**, 3823 (2012).
- [61] S. M. Loverde, D. A. Pantano, D. A. Christian, A. Mahmud, M. L. Klein, and D. E. Discher, *Curr. Opin. Solid State Mater. Sci.* **15**, 277 (2011).
- [62] J. R. Spaeth, I. G. Kevrekidis, and A. Z. Panagiotopoulos, *J. Chem. Phys.* **135**, 184903 (2011).
- [63] G. N. Condezo and C. San Martin, *PLoS Pathog.* **13** (2017).
- [64] G. D. Price and M. R. Badger, *Canadian Journal of Botany* **69**, 963 (1991).
- [65] J. M. Shively, F. Ball, D. H. Brown, and R. E. Saunders, *Science* **182**, 584 (1973).
- [66] J. M. Shively, F. L. Ball, and B. W. Kline, *J. Bacteriol.* **116**, 1405 (1973).
- [67] C. V. Iancu, D. M. Morris, Z. Dou, S. Heinhorst, G. C. Cannon, and G. J. Jensen, *J. Mol. Biol.* **396**, 105 (2010).
- [68] M. Sutter, T. G. Laughlin, N. B. Sloan, D. Serwas, K. M. Davies, and C. A. Kerfeld, *Plant Physiol.* **181**, 1050 (2019).

- [69] T. Chen and S. C. Glotzer, *Physical Review E* **75**, 051504 (2007).
- [70] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, *Phys. Rev. Lett.* **90**, 248101 (2003).
- [71] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15556 (2004).
- [72] B. Berger, P. W. Shor, L. Tuckerkellogg, and J. King, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7732 (1994).
- [73] R. Schwartz, R. L. Garcea, and B. Berger, *Virology* **268**, 461 (2000).
- [74] D. Rapaport, *Phys. Rev. E* **70**, 051905 (2004).
- [75] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, *J. Am. Chem. Soc.* **131**, 2606 (2009).
- [76] O. M. Elrad and M. F. Hagan, *Nano Lett.* **8**, 3850 (2008).
- [77] I. G. Johnston, A. A. Louis, and J. P. K. Doye, *J. Phys.: Condens. Matter* **22**, 104101 (2010).
- [78] D. C. Rapaport, *J. Phys.: Condens. Matter* **22**, 104115 (2010).
- [79] H. Nguyen and C. Brooks, *Nano Lett.* **8**, 4574 (2008).
- [80] M. G. Mateu, *Arch. Biochem. Biophys.* **531**, 65 (2013).
- [81] M. F. Hagan, *Adv. Chem. Phys.* **155**, 1 (2014).
- [82] J. D. Perlmutter and M. F. Hagan, *Annu. Rev. Phys. Chem.* **66**, 217 (2015), <http://dx.doi.org/10.1146/annurev-physchem-040214-121637>.
- [83] R. F. Bruinsma and W. S. Klug, *Annu. Rev. Condens. Matter Phys.* **6**, 245 (2015), <https://doi.org/10.1146/annurev-conmatphys-031214-014325>.
- [84] M. F. Hagan and R. Zandi, *Current Opinion in Virology* **18**, 36 (2016).
- [85] D. C. Rapaport, *Phys. Rev. E* **86**, 051917 (2012).
- [86] D. C. Rapaport, J. E. Johnson, and J. Skolnick, *Comput. Phys. Commun.* **122**, 231 (1999).
- [87] C. I. Mendoza and D. Reguera, *eLife* **9**, e52525 (2020).
- [88] J. Wagner and R. Zandi, *Biophys. J.* **109**, 956 (2015).
- [89] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger, *Biophys. J.* **75**, 2626 (1998).
- [90] M. F. Hagan and D. Chandler, *Biophys. J.* **91**, 42 (2006).
- [91] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, *J. Chem. Phys.* **127**, 085106 (2007).
- [92] J. E. Baschek, H. C. R. Klein, and U. S. Schwarz, *Bmc Biophysics* **5** (2012), 10.1186/2046-1682-5-22.
- [93] M. Castelnovo, T. Verdier, and L. Foret, *Epl* **105** (2014), 10.1209/0295-5075/105/28006.
- [94] M. Castelnovo, D. Muriaux, and C. Faivre-Moskalenko, *New J. Phys.* **15** (2013), 10.1088/1367-2630/15/3/035028, castelnovo, M. Muriaux, D. Faivre-Moskalenko, C.
- [95] M. A. Boettcher, H. C. R. Klein, and U. S. Schwarz, *Phys. Biol.* **12** (2015), 10.1088/1478-3975/12/1/016014.
- [96] S. N. Fejer, D. Chakrabarti, and D. J. Wales, *ACS Nano* **4**, 219 (2010).
- [97] R. Kusters, H.-K. Lin, R. Zandi, I. Tsvetkova, B. Dragnea, and P. van der Schoot, *J. Phys. Chem. B* **119**, 1869 (2015).
- [98] R. Zandi and P. van der Schoot, *Biophys. J.* **96**, 9 (2009).
- [99] M. F. Hagan, *Phys. Rev. E* **77**, 051904 (2008).
- [100] M. F. Hagan, *J. Chem. Phys.* **130**, 114902 (2009).
- [101] V. A. Belyi and M. Muthukumar, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17174 (2006).
- [102] R. Twarock, R. J. Bingham, E. C. Dykeman, and P. G. Stockley, *Curr. Opin. Vir.* **31**, 74 (2018).
- [103] R. Zandi, B. Dragnea, A. Travasset, and R. Podgornik, *Phys. Rep.* (2020).
- [104] S. Li, P. Roy, A. Travasset, and R. Zandi, *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10971 (2018).
- [105] S. Li, R. Zandi, A. Travasset, and G. M. Grason, *Phys. Rev. Lett.* (2019), 10.1103/PhysRevLett.123.145501, arXiv:1906.03301.
- [106] P. van der Schoot and R. Zandi, *J. Biol. Phys.* **39**, 289 (2013).
- [107] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, *Phys. Rev. E* **89**, 032707 (2014).
- [108] A. Siber, R. Zandi, and R. Podgornik, *Phys. Rev. E* **81**, 051919 (2010).
- [109] E. C. Dykeman, P. G. Stockley, and R. Twarock, *J. Mol. Biol.* **425**, 3235 (2013).
- [110] N. Patel, E. C. Dykeman, R. H. A. Coutts, G. P. Lomonosoff, D. J. Rowlands, S. E. V. Phillips, N. Ranson, R. Twarock, R. Tuma, and P. G. Stockley, *Proc. Natl. Acad. Sci. U.S.A.* (2015), 10.1073/pnas.1420812112.
- [111] E. C. Dykeman, P. G. Stockley, and R. Twarock, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5361 (2014).
- [112] E. C. Dykeman, P. G. Stockley, and R. Twarock, *Phys Rev E* **87**, 022717 (2013).
- [113] V. L. Morton, E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, and P. G. Stockley, *J. Mol. Biol.* **401**, 298 (2010).
- [114] R. Zhang, E. Wernersson, and P. Linse, *RSC Adv.* **3**, 25258 (2013).
- [115] R. Zhang and P. Linse, *J. Chem. Phys.* **140** (2014), 10.1063/1.4883056.
- [116] R. Zhang and P. Linse, *J. Chem. Phys.* **138** (2013), 10.1063/1.4799243.
- [117] D. G. Angelescu, J. Stenhammar, and P. Linse, *J. Phys. Chem. B* **111**, 8477 (2007).
- [118] D. G. Angelescu and P. Linse, *Phys. Rev. E* **75**, 051905 (2007).
- [119] J. P. Mahalik, K. A. Brown, X. Cheng, and M. Fuentes-Cabrera, *ACS Nano* **10**, 5751 (2016).
- [120] E. J. Young, R. Burton, J. P. Mahalik, B. G. Sumpter, M. Fuentes-Cabrera, C. A. Kerfeld, and D. C. Ducat, *Front. Microbiol.* **8**, 1441 (2017).
- [121] L. M. Oltrogge and D. F. Savage, “N.d.” (N.D.), unpublished.
- [122] A. Halperin, *Macromolecules* **20**, 2943 (1987).
- [123] M. Daoud and J. P. Cotton, *J. Phys.* **43**, 531 (1982).
- [124] R. Hariharan, C. Biver, J. Mays, and W. B. Russel, *Macromolecules* **31**, 7506 (1998).
- [125] A. Manandhar, M. Kang, K. Chakraborty, P. K. Tang, and S. M. Loverde, *Organic & biomolecular chemistry* **15**, 7993 (2017).
- [126] S. M. Lee, N. Bond, C. Callaway, B. Clark, E. Farmer, M. Mallard, and S. S. Jang, *RSC Adv.* **8**, 37866 (2018).
- [127] M. Sutter, M. Faulkner, C. Aussignargues, B. C. Paasch, S. Barrett, C. A. Kerfeld, and L.-N. Liu, *Nano Lett.* **16**, 1590 (2016), pMID: 26617073, <http://dx.doi.org/10.1021/acs.nanolett.5b04259>.
- [128] F. Cai, B. B. Menon, G. C. Cannon, K. J. Curry, J. M. Shively, and S. Heinhorst, *PLoS ONE* **4**, e7521 (2009).
- [129] P. Ryan, T. J. Forrester, C. Wroblewski, T. M. Kenney, E. N. Kitova, J. S. Klassen, and M. S. Kimber, *J. Biol. Chem.* **294**, 2593 (2019).
- [130] W. Dai, M. Chen, C. Myers, S. J. Ludtke, B. M. Pettitt, J. A. King, M. F. Schmid, and W. Chiu, *J. Mol. Biol.* **430**, 4156 (2018).
- [131] S. Alexander, *J. Phys.* **38**, 983 (1977).
- [132] T. A. Witten and P. A. Pincus, *Macromolecules* **19**, 2509 (1986).
- [133] E. B. Zhulina and O. V. Borisov, *Macromolecules* **31**, 7413 (1998).
- [134] M. F. Hagan and G. M. Grason, arXiv.2007.01927 (2020), <https://arxiv.org/abs/2007.01927>.
- [135] P. de Gennes, *Scaling concepts in polymer physics* (Cornell Univ. Pr., 1979) p. 324 S.

- [136] S. Heinhorst, G. C. Cannon, and J. M. Shively, in *Complex intracellular structures in prokaryotes* (Springer, 2006) pp. 141–165.
- [137] C. Fan, S. Cheng, Y. Liu, C. M. Escobar, C. S. Crowley, R. E. Jefferson, T. O. Yeates, and T. a. Bobik, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7509 (2010).
- [138] C. Aussignargues, B. C. Paasch, R. Gonzalez-Esquer, O. Erbilgin, and C. A. Kerfeld, *Communicative & integrative biology* **8**, e1039755 (2015).
- [139] J. A. Anderson, C. D. Lorenz, and A. Travesset, *J. Comput. Phys.* **227**, 5342 (2008).
- [140] T. D. Nguyen, C. L. Phillips, J. A. Anderson, and S. C. Glotzer, *Comput. Phys. Commun.* **182**, 2307 (2011).
- [141] .
- [142] M. Faulkner, J. Rodriguez-Ramos, G. F. Dykes, S. V. Owen, S. Casella, D. M. Simpson, R. J. Beynon, and L.-N. Liu, *Nanoscale* **9**, 10662 (2017).
- [143] J. Lidmar, L. Mirny, and D. R. Nelson, *Phys. Rev. E* **68** (2003).
- [144] T. T. Nguyen, R. F. Bruinsma, and W. M. Gelbart, *Phys. Rev. E* **72**, 051923 (2005).
- [145] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, *Nano Lett.* **7**, 338 (2007).
- [146] O. Elrad and M. F. Hagan, *Phys. Biol.* **7**, 045003 (2010).

SUPPLEMENTARY INFORMATION

S1. EQUILIBRIUM THEORY

Free energy. We consider shells composed of $n_{\text{shell}} = 4\pi R^2/a$ hexamer subunits, with R the shell radius and a the subunit area, and n_{scaf} scaffolds. We incorporate the presence of cargo implicitly through effective attractions between scaffold molecules embodied in the parameter $\Delta\mu$. The net free energy of an assembling shell with radius R is given by

$$\begin{aligned} \frac{F_{\text{assem}}}{k_B T} = & \frac{4\pi R^2}{a} \left[g_{\text{hh}} + \frac{\kappa}{2} \left(\frac{2}{R} - \frac{2}{R_0} \right)^2 \right] + \\ & n_{\text{scaf}} \left[\left(\frac{R}{R_{\text{scaf}}} \right)^2 + \left(\frac{R_{\text{scaf}}}{R} \right)^2 + \Delta\mu + \frac{\nu}{2} N_s C_{\text{scaf}} \right] + \\ & n_{\text{scaf}} \log \frac{n_{\text{scaf}}}{n_{\text{shell}}} + (n_{\text{shell}} - n_{\text{scaf}}) \log \frac{n_{\text{shell}} - n_{\text{scaf}}}{n_{\text{shell}}} \end{aligned} \quad (\text{S1})$$

The first term on the right-hand side of Eq. S1 represents the free energy due to subunit-subunit contacts that drive shell assembly, with g_{hh} per energy per subunit. The second term describes the elastic energy associated with deviations of the shell from its intrinsic spontaneous curvature, with κ the bending modulus. The next two terms give the entropic penalty for stretching or compressing scaffold molecules from their preferred length, R_{scaf} . In this model we have assumed that the scaffold bridges from the cargo in the interior to the shell surface and thus we set the scaffold length equal to the shell radius. The following two terms account for attractive scaffold-cargo and scaffold-shell interactions, and scaffold-scaffold excluded volume interactions, with the excluded volume per scaffold segment given by ν and the scaffold concentration in the shell given by

$C_{\text{scaf}} = \frac{n_{\text{scaf}} N_s}{4/3\pi R^3}$. The final two terms represent the mixing entropy of n_{scaf} scaffolds binding to a shell with n_{shell} subunits. For simplicity we have assumed that one scaffold can bind to each shell subunit, so the total number of binding sites is equal to n_{shell} .

By defining the surface density of the scaffolds $\sigma_s \equiv \frac{n_{\text{scaf}}}{n_{\text{shell}}}$, the total free energy per subunit $f_{\text{assem}} = \frac{F_{\text{assem}}}{4\pi R^2/a}$ of an assembled shell with radius R can be written as

$$\begin{aligned} f_{\text{assem}}(R(n), \sigma_s) = & g_{\text{hh}} + \frac{\kappa a}{2} \left(\frac{2}{R} - \frac{2}{R_0} \right)^2 + \\ & \sigma_s \left[\left(\frac{R}{R_{\text{scaf}}} \right)^2 + \left(\frac{R_{\text{scaf}}}{R} \right)^2 + \Delta\mu + \frac{3}{2} \frac{\nu \sigma_s N_s^2}{aR} \right] + \\ & \sigma_s \log \sigma_s + (1 - \sigma_s) \log(1 - \sigma_s) \end{aligned} \quad (\text{S2})$$

Optimal shell size. For all of the theoretical results presented in this work, we consider the case of limiting hexamer subunits, with the scaffold present in excess at fixed chemical potential μ_{scaf} , so that the term $\Delta\mu$ in Eq. S2 is a constant. The equilibrium distribution $\rho(n, \sigma_s)$ of shells with n hexamer subunits can then be calculated by variationally minimizing the total system free energy density

$$\begin{aligned} f_{\text{tot}}(\{\rho(n, \sigma_s)\}) = & \sum_{n, \sigma_s} \rho(n, \sigma_s) f_{\text{assem}}(R(n), \sigma_s) + \\ & \sum_{n, \sigma_s} \rho(n, \sigma_s) (\log \rho(n, \sigma_s) - 1) \end{aligned} \quad (\text{S3})$$

under the constraint of mass conservation: $\Phi = \sum_{n,\sigma_s}^{\infty} n\rho(n, \sigma_s)$ with Φ the total concentrations of hexamer subunits. This obtains a law of mass action

$$\rho(n, \sigma_s) = \exp(-[f_{\text{assem}}(R(n), \sigma_s) - \mu_h]n/k_B T) \quad (\text{S4})$$

with f_{assem} given by Eq. S2 and $\mu_h = k_B T \log(\rho_h v_0)$ the chemical potentials of free hexamer subunits at their equilibrium concentration ρ_h with v_0 a standard state volume. The optimal shell size n^* is then obtained by variationally minimizing Eq. S4 with respect to n and σ_s . The first of these operations results in

$$n^* \left. \frac{\partial f_{\text{assem}}}{\partial n} \right|_{n^*} + f_{\text{assem}}(R(n^*), \sigma_s) - \mu_h = 0. \quad (\text{S5})$$

It can be shown that above the pseudo-critical subunit concentration (i.e. the hexamer concentration above which there is significant assembly), the term $f_{\text{assem}}(n^*, \sigma_s) - \mu_h \approx 0$ in the limit $n^* \gg 1$ [134]. To see this, note that at equilibrium the chemical potential of free subunits μ_h must be equal to the chemical potential of a subunit in a shell, given by

$$\mu(n) = \frac{1}{n} \frac{\partial f_{\text{tot}}}{\partial \rho(n, \sigma_s) v_0} = f_{\text{assem}}(R(n), \sigma_s) + k_B T \frac{\log(\rho(n, \sigma_s) v_0)}{n} \quad (\text{S6})$$

with the last term in Eq. S6 accounting for the per-subunit free energy due to shell mixing entropy. In the limit of large shell size, the last term becomes negligible for shells at the highest concentrations (i.e. with $n = n^*$), resulting in $f_{\text{assem}}(R(n^*), \sigma_s) \approx \mu_h$.

Thus, the peak in the shell size distribution corresponds to the minimum free energy **per subunit**; i.e., $\left. \frac{\partial f_{\text{assem}}(R(n), \sigma_s)}{\partial n} \right|_{n^*} = 0$, or alternatively:

$$\begin{aligned} \frac{\partial f_{\text{assem}}}{\partial R} &= 0 \\ \frac{\partial f_{\text{assem}}}{\partial \sigma_s} &= 0. \end{aligned} \quad (\text{S7})$$

In Ref. 134 this calculation is performed rigorously, showing that for finite concentration the optimum shifts to smaller shell sizes, but that this shift is negligible for large n^* .

Finally, note that an analogous analysis could be performed for limiting scaffold molecules, in which case the peak assembly size would correspond to the minimum free energy per scaffold molecule.

Binding free energy of shell subunits. The per subunit shell binding free energy g_{hh} can be written as

$$g_{\text{hh}} = n_{\text{bonds}} \varepsilon_H \varepsilon_{\text{hh}} - T(s_H + s_{\text{config}}) \quad (\text{S8})$$

with $n_{\text{bonds}} = 3$ the number of bonds per subunit in the shell, s_{hex} the translational and rotational entropy penalty associated with binding of hexameric subunits and $s_{\text{config}} = -\log(6)$ accounting for the configurational entropy associated with the subunit's hexameric symmetry. $\varepsilon_H = -3.15$ and $s_H = -17.7$ are taken from Ref. 44, where they were measured from the dimerization equilibrium constant in simulations of subunits of a similar model only capable of forming dimers.

Approximations in the elastic energy. For simplicity we have modeled shell bending energy using the Helfrich energy for a fluid membrane. We note that for an elastic shell tiled by hexamer subunits, there are additional contributions arising from energy associated with the 12 five-fold disclinations subunits required by topology, and elastic interactions between these disclinations. However, as shown in Ref. 45, the free energy from the defects themselves simply renormalizes the intrinsic spontaneous curvature radius of the shell, with a favorable or unfavorable defect free energy respectively decreasing or increasing the effective spontaneous curvature radius. Since this does not qualitatively change the results, we have neglected this term for simplicity here. We note that while the defect energy in an elastic shell is typically positive (unfavorable), in the case of microcompartments it may be positive or negative since the disclinations are filled by pentamer subunits, which have a different protein sequence (and thus potentially different binding affinity and different preferred binding angles) from the hexamer and pseudo-hexamer proteins that tile the rest of the shell.

The nature of the inter-defect interactions depends on the dimensionless Foppl von Karmann number ($\gamma \equiv YR^2/\kappa$ for a shell of radius R with Young's modulus Y and bending modulus κ). Below a critical value of the Foppl von Karmann number ($\gamma = 154$), the shells remain roughly spherical [143, 144] and the energy of the inter-disclination elastic interactions is proportional to the shell area. Thus, its contribution can be subsumed into the hexamer-hexamer interaction g_{hh} (whose contribution also is linear in shell area) [45]. Above the critical Foppl von Karmann number the shells facet and the inter-disclination energy is screened [143]. Thus, the form of the Helfrich energy is sufficient to describe the equilibrium bending energy of the spherical shells that we consider. However, we note that the inter-defect interactions can have a significant effect on the energy landscape and preferred geometry of assembly intermediates [104, 105], and thus would be important to consider for assembly pathways and kinetics.

C. Scaling analysis

Here we present a more detailed derivation of the asymptotic results presented in the main text.

1. Optimal shell size

A. Low intrinsic shell curvature, $R_0 \gtrsim R_E$. In this limit the contribution of the intrinsic shell curvature can be neglected, so the shell and scaffold free energy terms in Eq. 1 of the main text simplify to

$$f_{\text{shell}} = \left(\frac{R_E}{R} \right)^2 \quad (\text{S9})$$

$$f_{\text{scaf}} = \left(\frac{R}{R_{\text{scaf}}} \right)^2 + \left(\frac{R_{\text{scaf}}}{R} \right)^2 + \Delta\mu. \quad (\text{S10})$$

Minimizing the free energy at fixed scaffold incorporation σ then results in an optimal shell size given by Eq. 6 of the main text.

B. High intrinsic shell curvature, $R_0 < R_E$. In this regime the preferred curvature of the shell proteins influences the assembled shell size, so we must consider the full free energy Eq. S2 except we retain the approximation $\nu = 0$. As above, we consider separately the limits in which the preferred scaffold size is larger or smaller than the characteristic shell size.

(1) $R_{\text{scaf}} \gg R_0$: In this case we simplify the scaffold confinement free energy by neglecting the stretching term:

$$f_{\text{shell}} = R_E^2 \left(\frac{1}{R} - \frac{1}{R_0} \right)^2 \quad (\text{S11})$$

$$f_{\text{scaf}} = \left(\frac{R_{\text{scaf}}}{R} \right)^2 + \Delta\mu \quad (\text{S12})$$

and minimizing the free energy results in

$$R_{\text{eq}} \approx R_0 \left[1 + \sigma_{\text{eq}} (R_{\text{scaf}}/R_E)^2 \right]. \quad (\text{S13})$$

(2) $R_{\text{scaf}} \ll R_0$: In this limit we retain the scaffold stretching free energy term:

$$f_{\text{shell}} = R_E^2 \left(\frac{1}{R} - \frac{1}{R_0} \right)^2 \quad (\text{S14})$$

$$f_{\text{scaf}} = \left(\frac{R}{R_{\text{scaf}}} \right)^2 + \Delta\mu. \quad (\text{S15})$$

Recognizing that the shell size will remain close to the shell protein spontaneous curvature radius, we set

$$R_{\text{eq}} \approx R_0(1 - \epsilon). \quad (\text{S16})$$

Inserting this into Eq. S16, minimizing with respect to ϵ and σ , and expanding to linear order in ϵ results in

$$\epsilon^{-1} = 3 + \frac{R_{\text{scaf}}^2 R_E^2}{\sigma R_0^4} \quad (\text{S17})$$

and an optimal shell size

$$\begin{aligned} R_{\text{eq}} &= R_0(2\sigma R_0^4 + R_{\text{scaf}}^2 R_E^2)/(3\sigma R_0^4 + R_{\text{scaf}}^2 R_E^2) \\ &\approx R_0 \left(1 - \frac{\sigma R_0^4}{R_{\text{scaf}}^2 R_E^2} \right) \end{aligned} \quad (\text{S18})$$

2. Maximum and minimum optimal shell sizes

Here we present details on estimating the maximum and minimum shell sizes that lead to significant scaffold packaging.

A. Small intrinsic shell curvature, $R_0 \gtrsim R_E$. In this limit the scaffold confinement free energy dominates over shell bending energy, and thus we assume the confinement free energy is approximately given by its minimum value $f_{\text{scaf}} \approx \Delta\mu + 2k_B T$. Scaffold packaging is then determined by a competition between the free energy driving scaffold packaging $\Delta\mu$ and the shell bending energy:

$$f_{\text{shell}} = R_E^2 \left(\frac{1}{R} - \frac{1}{R_0} \right)^2. \quad (\text{S19})$$

Including the scaffold confinement and mixing free energy terms results in an effective free energy per shell subunit given by:

$$\Delta\mu_{\text{eff}} \equiv -(\Delta\mu + 2k_B T)\sigma_{\text{eq}} - f_{\text{ent}} \quad (\text{S20})$$

Setting $\Delta\mu_{\text{eff}} = f_{\text{shell}}$ results in two solutions corresponding to Eqs. (8) and (9) of the main text.

B. High intrinsic shell curvature, $R_0 < R_E$. As shown above, the shell size remains close to R_0 in this limit, so scaffold packaging depends on a competition between its confinement free energy f_{scaf} and its attractive scaffold-cargo and scaffold-shell interactions represented by $\Delta\mu$. The nature of this competition depends on whether the preferred scaffold size is larger or smaller than the shell size as follows.

(1) $R_{\text{scaf}} > R_0$. In this case we focus on the scaffold confinement free energy term, and balance this against $\Delta\mu$:

$$\left(\frac{R_{\text{scaf}}}{R_{\text{eq}}} \right)^2 + 1 \approx \left[R_0^2 \left(\frac{1}{R_{\text{scaf}}^2} + \frac{2\sigma_{\text{eq}}}{R_E^2} \right) \right]^{-1} + 1 = |\Delta\mu| \quad (\text{S21})$$

where the +1 on the left hand side of Eq. S21 accounts for the scaffold stretching term. Rearranging to solve for R_{scaf} results in the maximum shell size $R_{\text{eq}}^{\text{max}}$ given in Eq. 12 of the main text.

(1) $R_{\text{scaf}} < R_0$. Here we balance the scaffold stretching term against $\Delta\mu$:

$$\begin{aligned} \left(\frac{R_{\text{eq}}}{R_{\text{scaf}}} \right)^2 + 1 &\approx \left(\frac{R_0^2}{R_{\text{scaf}}} \right)^2 \left(1 - \frac{\sigma_{\text{eq}} R_0^4}{R_{\text{scaf}}^2 R_E^2} \right)^2 + 1 \\ &\approx \left(\frac{R_0^2}{R_{\text{scaf}}} \right)^2 \left(1 - \frac{2\sigma_{\text{eq}} R_0^4}{R_{\text{scaf}}^2 R_E^2} \right) + 1 = |\Delta\mu| \end{aligned} \quad (\text{S22})$$

Solving for R_{scaf} and taking the limit $R_0 \ll R_E$ results in $R_{\text{eq}}^{\text{min}}$ given in Eq. 12 of the main text.

S2. COMPUTATIONAL MODEL DETAILS

Our model represents subunits as rigid bodies comprised of pseudoatoms arranged to capture the directional attractions and shape of microcompartment pentamer and hexamer oligomers. In comparison to earlier studies with patchy spheres (e.g. [89–92]), multi-pseudoatom subunits better describe the subunit excluded volume shape [74, 145, 146], which we find to be important for representing assembly around many-molecule cargoes. See Ref. 81 for a comparison of these approaches.

Scaffold molecules. Scaffolds are modeled as flexible bead-spring polymers with three domains; cargo-binding beads are at one end, shell-binding beads are at the other end, and beads in the middle domain have only repulsive interactions with cargo molecules and shell subunits. There are weak attractive interactions between pairs of scaffold beads.

The attractive scaffold-cargo interactions lead to collapse of the scaffold-cargo domain, resulting in a shorter effective length of encapsulated scaffolds relative to their free end-to-end size. Therefore, to compare simulation results against the theoretical model (in Fig. 3)), we estimated the effective length of the scaffold-cargo domain that would result in a comparable size, resulting in an effective length of $L_{\text{sc}}^{\text{eff}} = 2.5$ for the actual value $L_{\text{sc}} = 7$ used in the simulations.

Interaction potentials

In our model, all potentials can be decomposed into pairwise interactions. Potentials involving shell subunits further decompose into pairwise interactions between their constituent building blocks – the excluders, attractors, ‘Top’, and ‘Bottom’

pseudoatoms. It is convenient to write the total energy of the system as the sum of 6 terms:, involving shell-shell (U_{hh}), scaffold-shell (U_{sh}), scaffold-cargo (U_{sc}), scaffold-scaffold (U_{ss}), cargo-cargo (U_{cc}) and shell-cargo (U_{hc}) interactions, each summed over all pairs of the appropriate type:

$$U = \sum_{\text{shell } i} \sum_{\text{shell } j < i} U_{hh} + \sum_{\text{scaffold } i} \sum_{\text{shell } j} U_{sh} + \sum_{\text{scaffold } i} \sum_{\text{cargo } j} U_{sc} + \sum_{\text{scaffold } i} \sum_{\text{scaffold } j < i} U_{ss} + \sum_{\text{cargo } i} \sum_{\text{cargo } j < i} U_{cc} + \sum_{\text{shell } i} \sum_{\text{cargo } j} U_{hc} \quad (\text{S23})$$

where $\sum_{\text{shell } i} \sum_{\text{shell } j < i}$ is the sum over all distinct pairs of shell subunits in the system, $\sum_{\text{scaffold } i} \sum_{\text{cargo } j}$ is the sum over all shell-scaffold particle pairs, etc.

Shell-shell interaction potentials. The shell-shell potential U_{ss} is the sum of the attractive interactions between complementary attractors, and geometry guiding repulsive interactions between ‘Top’ - ‘Top’, ‘Bottom’ - ‘Bottom’, and ‘Top’ - ‘Bottom’ pairs. There are no interactions between members of the same rigid body. For notational clarity, we index rigid bodies and non-rigid pseudoatoms in Roman, while the pseudoatoms comprising a particular rigid body are indexed in Greek. For subunit i we denote its attractor positions as $\{\mathbf{a}_{i\alpha}\}$ with the set comprising all attractors α , its ‘Top’ position $\mathbf{t}_i$, ‘Bottom’ position $\mathbf{b}_i$ and the ‘M’ pseudoatom at the center of the subunit in the plane of the attractors, as $\mathbf{m}_i$.

The shell-shell interaction potential between two subunits i and j is then defined as:

$$U_{hh}(\{\mathbf{a}_{i\alpha}\}, \mathbf{t}_i, \mathbf{b}_i, \mathbf{m}_i, \{\mathbf{a}_{j\beta}\}, \mathbf{t}_j, \mathbf{b}_j, \mathbf{m}_j) = \varepsilon_{\text{angle}} \mathcal{L}(|\mathbf{t}_i - \mathbf{t}_j|, \sigma_{t,ij}) + \varepsilon_{\text{angle}} \mathcal{L}(|\mathbf{b}_i - \mathbf{b}_j|, \sigma_b) + \varepsilon_{\text{angle}} \mathcal{L}(|\mathbf{b}_i - \mathbf{t}_j|, \sigma_{tb}) + \mathcal{L}(|\mathbf{m}_i - \mathbf{m}_j|, \sigma_m) + \sum_{\alpha, \beta}^{N_{ai}, N_{aj}} \varepsilon_{hh} \mathcal{M}(|\mathbf{a}_{i\alpha} - \mathbf{a}_{j\beta}|, r_0, \varrho, r_{\text{cut}}^{\text{att}}) \quad (\text{S24})$$

The function $\mathcal{L}$ is defined as the repulsive component of the Lennard-Jones potential shifted to zero at the interaction diameter:

$$\mathcal{L}(x, \sigma) \equiv \theta(\sigma - x) \left[\left(\frac{\sigma}{x} \right)^{12} - 1 \right] \quad (\text{S25})$$

with $\theta(x)$ the Heaviside function. The function $\mathcal{M}$ is a Morse potential:

$$\mathcal{M}(x, r_0, \varrho, r_{\text{cut}}) = \theta(r_{\text{cut}} - x) \times \left[\left(e^{\left(1 - \frac{x}{r_0} \right)} - 2 \right) e^{\left(1 - \frac{x}{r_0} \right)} - V_{\text{shift}}(r_{\text{cut}}) \right] \quad (\text{S26})$$

with $V_{\text{shift}}(r_{\text{cut}})$ the value of the (unshifted) potential at r_{cut} .

The parameter ε_{hh} sets the strength of the shell-shell attraction at each attractor site, N_{ai} is the number of attractor pseudoatoms in subunit i , and $\varepsilon_{\text{angle}}$ scales the repulsive interactions that enforce the geometry.

Shell-shell interaction parameter values. *Attractors:* The strength of attractive interactions is parameterized by the well-depth ε_{hh} for a pair of attractors on hexamers as follows. Hexamer-Hexamer edge attractor pairs have a well-depth of ε_{hh} . Because vertex attractors have multiple partners in an assembled structure, whereas edge attractors have only one, the well-depth for the vertex pairs is set to $0.5\varepsilon_{hh}$. The parameter r_0 is the minimum energy attractor distance, set to 0.2, $\varrho = 5$ determines the width of the attractive interaction, and $r_{\text{cut}}^{\text{att}} = 2.0$ is the cutoff distance for the attractor potential.

Repulsive interactions and simulated shell bending modulus: The ‘Top’ and ‘Bottom’ heights, or distance out of the attractor plane, are set to $h = 1/2$. The diameter of the ‘Top’ - ‘Bottom’ interaction, which prevents subunits from binding in inverted configurations [77], is $\sigma_{tb} = 1.8$. A central excluder ‘M’ with effective diameter $\sigma_m = 2.026$ in the center of subunits prevents subunit overlaps. The effective diameters of the ‘Bottom’ - ‘Bottom’ interaction σ_b and the ‘Top’ - ‘Top’ interaction σ_t determine the shell spontaneous curvature radius as follows: $R_0 = 22$, $\sigma_b = 1.6$, $\sigma_t = 2.2$; $R_0 = 8.0$, $\sigma_b = 1.5$, $\sigma_t = 2.3$; and $R_0 = 4.5$, $\sigma_b = 1.4$, $\sigma_t = 2.4$.

The bending modulus of α -carboxysomes has not been experimentally estimated, but measurements on β -carboxysome shells obtained $\kappa \approx 25k_B T$ [142], which is thus smaller than that for a typical viral capsid. As noted above, our simulated hexamer subunits, with diameter 13 nm, are larger than the hexamer size (7 nm diameter) in a carboxysome. To maintain the characteristic

elastic length scale $R_E = \sqrt{2ak}$ of the simulated shells approximately equal to that of β -carboxysome shells $R_E \approx 47$ nm, despite the large size of hexamer subunits, we set $\varepsilon_{\text{angle}} = 0.5k_B T$, which corresponds to a bending modulus of about 5-10 $k_B T$ [45], and thus $R_E \approx 38 - 54$ nm.

Scaffold-shell interactions. The scaffold-shell interaction is modeled by a short-range repulsion between pairs of scaffold beads and scaffold-excluder pseudoatoms on subunits representing the excluded volume, plus an attractive interaction between pairs of shell-interacting beads in scaffolds and hexamer ‘Bottom’ pseudoatoms. For subunit i with excluder positions $\{\mathbf{x}_{i\alpha}\}$ and ‘Bottom’ pseudoatom position $\mathbf{b}_i$, and scaffold j with bead positions $\mathbf{s}_{j\beta}$ and shell-interacting beads $j\gamma$ with positions $\mathbf{s}_{j\gamma}$, the potential is:

$$U_{\text{sh}}(\{\mathbf{x}_{i\alpha}\}, \mathbf{b}_i, \{\mathbf{s}_{j\beta}\}, \{\mathbf{s}_{j\gamma}\}) = \sum_{\beta}^{N_s} \sum_{\alpha}^{N_x} \mathcal{L}(|\mathbf{x}_{i\alpha} - \mathbf{s}_{j\beta}|, \sigma_{\text{exs}}) + \sum_{\gamma}^{N_{\text{sh}}} \varepsilon_{\text{sh}} \mathcal{M}(|\mathbf{b}_i - \mathbf{s}_{j\gamma}|, r_0, \varrho^{\text{sh}}, r_{\text{cut}}^{\text{sh}}) \quad (\text{S27})$$

where ε_{sh} parameterizes the scaffold-shell interaction strength, N_s and N_{sh} are the total number of beads and the number of shell-interacting beads in each scaffold, N_x is the numbers of excluders on a shell subunit, $\sigma_{\text{exs}} = 0.375$ and $\sigma_t = 0.375$ are the effective diameters of the excluder - scaffold repulsions, $r_0^{\text{sh}} = 0.375$ is the minimum energy attractor distance, the width parameter is $\varrho^{\text{sh}} = 2.5$, and the cutoff is set to $r_{\text{cut}}^{\text{sh}} = 3.0$.

Scaffold-cargo interactions. The interaction between cargo particles and scaffolds is given by

$$U_{\text{sc}}(\mathbf{c}_i, \{\mathbf{s}_{j\alpha}\}, \{\mathbf{s}_{j\beta}\}) = \sum_{\alpha}^{N_s} \mathcal{L}(|\mathbf{c}_i - \mathbf{s}_{j\alpha}|, \sigma_{\text{sc}}) + \sum_{\beta}^{N_{\text{sc}}} \varepsilon_{\text{sc}} \mathcal{L}(|\mathbf{c}_i - \mathbf{s}_{j\beta}|, \sigma_{\text{sc}}, r_{\text{cut}}^{\text{sc}}) \quad (\text{S28})$$

with $\mathcal{L}$ the full Lennard-Jones interaction:

$$\mathcal{L}(x, \sigma, r_{\text{cut}}) = \theta(x - r_{\text{cut}}) \times \left\{ 4 \left[\left(\frac{x}{\sigma} \right)^{12} - \left(\frac{x}{\sigma} \right)^6 \right] - V_{\text{shift}}(r_{\text{cut}}) \right\} \quad (\text{S29})$$

with ε_{sc} an adjustable parameter that sets the strength of the scaffold-cargo interaction, N_s as the total number of scaffold beads, N_{sc} as the number of beads in the scaffold-cargo binding domain, $\sigma_{\text{sc}} = 0.75$ as the effective cargo-scaffold excluded volume size, and $r_{\text{cut}}^{\text{sc}} = 3\sigma_{\text{sc}}$ the interaction cut off length.

Scaffold-scaffold interactions. The scaffold-scaffold non-bonded interaction is modeled by a Lennard-Jones potential; in addition, segments occupying adjacent positions along the polymer chain interact through harmonic bonds.

$$U_{\text{ss}}(\mathbf{R}_i, \mathbf{R}_j) = \begin{aligned} &K_{\text{bond}}(\mathbf{R}_{ij}, \sigma_s, \kappa) : \{i, j\} \text{ bonded} \\ &\varepsilon_{\text{ss}} \mathcal{L}(\mathbf{R}_{ij}, \sigma_{\text{ss}}, r_{\text{cut}}^{\text{ss}}) : \{i, j\} \text{ nonbonded} \end{aligned} \quad (\text{S30})$$

where $R_{ij} \equiv |\mathbf{R}_i - \mathbf{R}_j|$ is the center-to-center distance between the scaffold beads, ε_{ss} is the well-depth of the Lennard-Jones (Eq. S29) potential, $\sigma_{\text{ss}} = 0.5$ is the effective scaffold-scaffold excluded volume size, and $r_{\text{cut}}^{\text{ss}} = 3\sigma_{\text{ss}}$ the interaction cut off length.

Bonds are represented by a harmonic potential:

$$K_{\text{bond}}(\mathbf{R}_{ij}, \sigma, \kappa) \equiv \frac{k_{\text{bond}}}{2} (R_{ij} - \sigma)^2 \quad (\text{S31})$$

Cargo-cargo interactions. The cargo-cargo interaction is modeled by a short-range repulsion interaction representing the excluded volume of cargo particles, given by:

$$U_{cc}(\mathbf{c}_i, \mathbf{c}_j) = \mathcal{L}(|\mathbf{c}_i - \mathbf{c}_j|, \sigma_c) \quad (\text{S32})$$

Shell-cargo interactions. The shell-cargo interaction is modeled by a short-range repulsion between pairs of cargo particles and cargo-excluder pseudoatoms on subunits representing the excluded volume:

$$U_{hc}(\{\mathbf{x}_{i\alpha}\}, \mathbf{c}_j) = \sum_{\alpha} \mathcal{L}(|\mathbf{x}_{i\alpha} - \mathbf{c}_j|, \sigma_{exc}) \quad (\text{S33})$$

FIG. S1. Equilibrium theory prediction for the amount of packaged scaffold σ_s as a function of R_0 and R_{scaf} . Predictions are obtained by numerically minimizing Eq. 1 with $R_E = 30r_*$, and $\nu = 0$. Results are shown for (A) $\Delta\mu = -10$ and (B) $\Delta\mu = -2.5$. (C) Minimum and maximum shell radius (normalized by the shell elastic energy length scale R_E) in the high intrinsic shell curvature regime ($R_0 < R_E$). The maximum and minimum shell radii that can form for varying scaffold length are shown, calculated by numerically minimizing Eq. 1 (symbols) and from the asymptotic analysis Eq. 11 (lines) at different scaffold binding affinities $\Delta\mu$ for shell spontaneous curvature $R_0 = 8$, $R_E = 30$ and $\nu = 0$.

FIG. S2. (A) End-to-end distance (R_{scaf}) of scaffolds encapsulated in shells as a function of scaffold length. For comparison, the unperturbed scaffold sizes (from panel (B)) are shown as gray 'x' symbols and the dashed line. (B) End-to-end distance of isolated scaffold molecules (i.e., free in solution rather than packaged in a shell) as a function of number of segments N_s , with the scaffold contour length given by $L_s = l_k N_s$ with $l_k = 0.5r_*$. The plot shows the end-to-end length measured with weak attractions between scaffold segments, $\epsilon_{ss} = 0.2$ (as used in the assembly simulations), and for comparison, the end-to-end scaling with purely repulsive inter-segment interactions.

FIG. S3. (A) Dependence of the number of encapsulated cargo particles on the scaffold length. (B) Dependence of the volume fraction of encapsulated particles (cargo and scaffolds) on the scaffold length. (C) Dependence of the shell size on the scaffold cargo-binding domain fraction $f_{sc} = L_{sc}/L_s$, for $R_0 = 22r_*$, $L_s = 64r_*$, and $L_{sh} = 7r_*$. Other parameter values for (A), (B) and (C) are as follows. Shell subunit-subunit affinities: $\epsilon_{hh} = 3.15$ at $R_0 = 4.5$, $\epsilon_{hh} = 2.85$ at $R_0 = 8.0$ and $\epsilon_{hh} = 2.65$ at $R_0 = 22$; scaffold-shell affinity $\epsilon_{sh} = 2.5$, and scaffold-cargo affinity $\epsilon_{sc} = 1.0$.

FIG. S4. Distribution of shell sizes at three scaffold lengths and three different radii of curvature of Fig. 3.

FIG. S5. (A) Dependence of the mean shell radius on the scaffold-shell and shell-shell affinities. The scaffold length is $L_s = 24$ ($L_{sc} = 7$, $L_{sm} = 10$), the shell spontaneous curvature radius is $R_0 = 8.0r_*$ and the scaffold-cargo affinity is $\epsilon_{sc} = 1k_B T$. Note that in the large L_s limit where relatively few scaffold molecules are packaged, strong shell-shell interactions ($\epsilon_{ss} \gtrsim 3.15$ at $R_0 = 4.5$, $\epsilon_{ss} \gtrsim 2.85$ at $R_0 = 8$ and $\epsilon_{ss} \gtrsim 2.65$ at $R_0 = 22$) are required for complete shell assembly. (B) Dependence of the mean shell radius on shell-subunit and scaffold concentrations. The scaffold length is $L_s = 64$ ($L_{sc} = 7$, $L_{sm} = 50$), the shell spontaneous curvature radius is $R_0 = 8.0r_*$, and the scaffold-cargo affinity is $\epsilon_{sc} = 1.0k_B T$. In (A) and (B), the parameter values corresponding to the simulation results of Figs. (2) and (3) are marked on the plot by 'x' symbols.