# Supporting Information - Mechanisms of Scaffold-Mediated Microcompartment Assembly and Size-Control

Farzaneh Mohajerani,[†] Evan Sayer,[†] Christopher Neil,[‡] Koe Inlow,[‡] and Michael F. Hagan[∗,†]

†Martin A. Fisher School of Physics, Brandeis University, Waltham, MA 02453, USA

‡Department of Biochemistry, Brandeis University, Waltham, MA 02453, USA

E-mail: hagan@brandeis.edu

## S1. Equilibrium Theory

***Free energy.*** We consider shells composed of $n_{\text{shell}} = 4\pi R^2/a$ shell subunits, with $R$ the shell radius and $a$ the subunit area, and $n_{\text{scaf}}$ 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

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

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 $g_{hh}$ the energy per subunit. The second term describes the elastic energy associated with deviations of the shell from its intrinsic spontaneous curvature, with $\kappa$ 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 $v$ and the scaffold concentration in the shell given by $C_{scaf} = \frac{n_{scaf}N_s}{4/3\pi R^3}$. 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 $\sigma_s \equiv \frac{n_{scaf}}{n_{shell}}$, the total free energy per subunit $f_{assem} = \frac{F_{assem}}{4\pi R^2/a}$ of an assembled shell with radius $R$ can be written as

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

***Optimal shell size.*** For all the theoretical results presented in this work, we consider the case of limiting shell subunits, with the scaffold present in excess at fixed chemical potential $\mu_{scaf}$, so that the term $\Delta\mu$ in Eq. (S2) is a constant. The equilibrium distribution $\rho(n, \sigma_s)$ of shells with $n$

subunits can then be calculated by variationally minimizing the total system free energy density

$$f_{\text{tot}}(\{\rho(n, \sigma_s)\}) = \sum_{n,\sigma_s} \rho(n, \sigma_s) f_{\text{assem}}(R(n), \sigma_s) +$$

$$k_B T \sum_{n,\sigma_s} \rho(n, \sigma_s) \left( \log \rho(n, \sigma_s) - 1 \right) \tag{S3}$$

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

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

with $f_{\text{assem}}$ given by Eq. (S2) and $\mu_h = k_B T \log(\rho_h v_0)$ the chemical potentials of free subunits at their equilibrium concentration $\rho_h$ 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 $\sigma_s$. The first of these operations results in

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

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

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

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 sub-**

**unit**; i.e., $\frac{\partial f_{\text{assem}}(R(n),\sigma_s)}{\partial n}\Big|_{n^*} = 0$, or alternatively:

$$\frac{\partial f_{\text{assem}}}{\partial R} = 0$$
$$\frac{\partial f_{\text{assem}}}{\partial \sigma_s} = 0. \tag{S7}$$

In Ref. 1 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 $g_{\text{hh}}$ can be written as

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

with $n_{\text{bonds}} = 3$ the number of bonds per subunit in the shell, $s_{\text{hex}}$ the translational and rotational entropy penalty for subunit binding and $s_{\text{config}} = -\log(6)$ accounting for the configurational entropy associated with the subunit's six-fold symmetry. $\varepsilon_H = -3.15$ and $s_H = -17.7$ are taken from Ref. 2 , 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 the shell bending energy by the Helfrich energy for a fluid membrane. 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. 3, 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 Karman (FvK) number, $\gamma \equiv YR^2/\kappa$ for a shell of radius R with Young's modulus Y and bending modulus $\kappa$. Below a critical value of the FvK number ($\gamma_{\mathrm{c}} = 154$), the shells remain roughly spherical[4,5] and the energy of the inter-disclination elastic interactions is proportional to the shell area. Thus, its contribution can be subsumed into the subunit-subunit interaction $g_{\mathrm{hh}}$ (whose contribution also is linear in shell area).[3] Above the critical value $\gamma_{\mathrm{c}}$ the shells facet and the inter-disclination energy is screened.[4] 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,[6,7] and thus would be important to consider for assembly pathways and kinetics.

## Scaling analysis

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

**Optimal shell size**

*A. **Low intrinsic shell curvature**, $R_0 \gtrsim R_{\mathrm{E}}$.* 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

$$\frac{f_{\mathrm{shell}}}{k_{\mathrm{B}}T} = \left(\frac{R_{\mathrm{E}}}{R}\right)^2 \tag{S9}$$

$$\frac{f_{\mathrm{scaf}}}{k_{\mathrm{B}}T} = \left(\frac{R}{R_{\mathrm{scaf}}}\right)^2 + \left(\frac{R_{\mathrm{scaf}}}{R}\right)^2 + \frac{\Delta\mu}{k_{\mathrm{B}}T}. \tag{S10}$$

5

Minimizing the free energy at fixed scaffold incorporation $\sigma$ then results in an optimal shell size given by Eq. 1 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) $R_{scaf} \gg R_0$*: In this case we simplify the scaffold confinement free energy by neglecting the stretching term:

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

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

and minimizing the free energy results in

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

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

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

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

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

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

Inserting this into Eq. (S16), minimizing with respect to $\epsilon$ and $\sigma$, and expanding to linear order in

$\epsilon$ results in

$$\epsilon^{-1} = 3 + \frac{R_{\text{scaf}}^2 R_{\text{E}}^2}{\sigma R_0^4} \tag{S17}$$

and an optimal shell size

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

.

## 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*, $R_0 \gtrsim R_{\text{E}}$. 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 $f_{\text{scaf}} \approx \Delta\mu + 2k_{\text{B}}T$. Scaffold packaging is then determined by a competition between the free energy driving scaffold packaging $\Delta\mu$ and the shell bending energy:

$$\frac{f_{\text{shell}}}{k_{\text{B}}T} = R_{\text{E}}^2 \left(\frac{1}{R} - \frac{1}{R_0}\right)^2. \tag{S19}$$

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

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

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

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

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 $R_{eq}^{max}$ given in Eq. 12 of the main text.

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

$$\left(\frac{R_{eq}}{R_{scaf}}\right)^2 + 1 \approx \left(\frac{R_0^2}{R_{scaf}^2}\right)^2\left(1 - \frac{\sigma_{eq}R_0^4}{R_{scaf}^2 R_E^2}\right)^2 + 1 \tag{S22}$$

$$\approx \left(\frac{R_0^2}{R_{scaf}^2}\right)^2\left(1 - \frac{2\sigma_{eq}R_0^4}{R_{scaf}^2 R_E^2}\right) + 1 = \frac{|\Delta\mu|}{k_B T} \tag{S23}$$

Solving for $R_{scaf}$ and taking the limit $R_0 \ll R_E$ results in $R_{eq}^{min}$ 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 hexameric oligomers. In comparison to earlier studies with patchy spheres (e.g.[8–11]), multi-pseudoatom subunits better describe the subunit excluded volume shape,[12–14] which we find to be important for representing assembly around
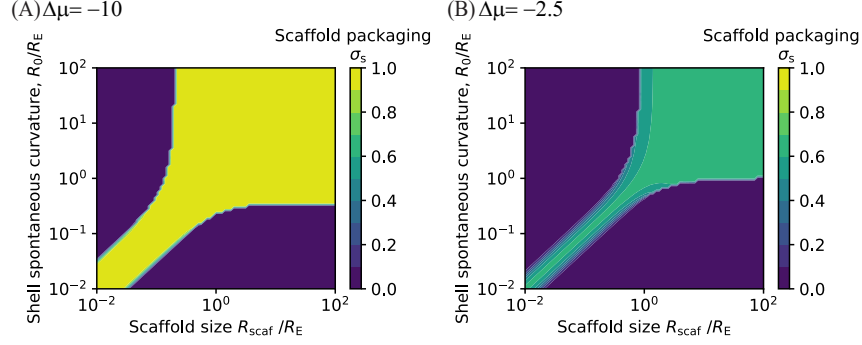
Figure S1: Equilibrium theory prediction for the amount of packaged scaffold $\sigma_s$ as a function of $R_0$ and $R_{\text{scaf}}$. Predictions are obtained by numerically minimizing Eq. 1 with $R_E = 30$, and $\nu = 0$. Results are shown for **(A)** $\Delta\mu = -10$ and **(b)** $\Delta\mu = -2.5$.

many-molecule cargoes. See Ref. 15 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.
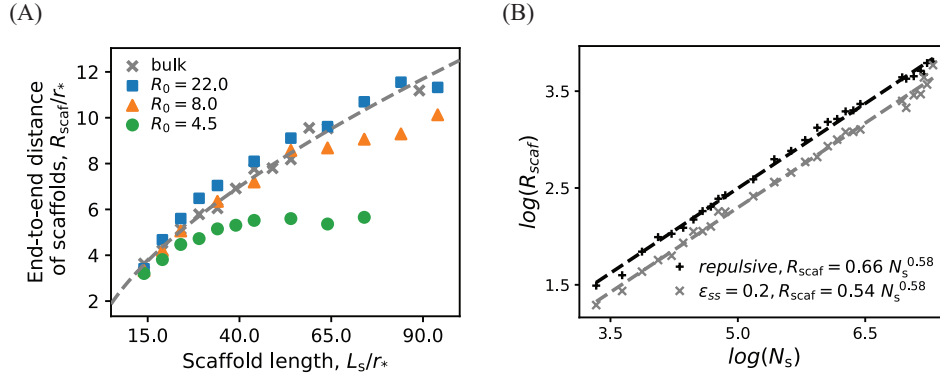


Figure S2: **(A)** End-to-end distance ($R_{\text{scaf}}$) of scaffolds encapsulated in shells as a function of scaffold length. For comparison, the unperturbed scaffold sizes (from panel (B)) are shown as gray 'x' symbols and a dashed line. **(B)** End-to-end distance of isolated scaffold molecules (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 $l_k = 0.5 r_*$. The plot shows the end-to-end length measured with weak attractions between scaffold segments, $\varepsilon_{ss} = 0.2$ (as used in the assembly simulations), and for comparison, the end-to-end scaling with purely repulsive inter-segment interactions.

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 $L_{sc}^{eff} = 2.5$ for the actual value $L_{sc} = 7$ 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 interactons 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 subunit-subunit ($U_{hh}$), 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:

$$U = \sum_{\text{shell } i} \sum_{\text{shell } j<i} U_{hh} + \sum_{\text{scaffold } i} \sum_{\text{shell } j} U_{sh} +$$
$$\sum_{\text{scaffold } i} \sum_{\text{cargo } j} U_{sc} + \sum_{\text{scaffold } i} \sum_{\text{scaffold } j<i} U_{ss} +$$
$$\sum_{\text{cargo } i} \sum_{\text{cargo } j<i} U_{cc} + \sum_{\text{shell } i} \sum_{\text{cargo } j} U_{hc} \tag{S24}$$

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 subunit-scaffold particle pairs, *etc.*

**Subunit-subunit interaction potentials.** The subunit-subunit potential $U_{ss}$ is the sum of the attractive interactions between complementary attractors, and geometry guiding repulsive inter-actions 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 $\alpha$, 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 subunit-subunit interaction potential between two subunits $i$ and $j$ is then defined as:

$$U_{\text{hh}}(\{\mathbf{a}_{i\alpha}\}, \mathbf{t}_i, \mathbf{b}_i, \mathbf{m}_i, \{\mathbf{a}_{j\beta}\}, \mathbf{t}_j, \mathbf{b}_j, \mathbf{m}_j) = \tag{S25}$$

$$\varepsilon_{\text{angle}} \text{Ł}\left(\left|\mathbf{t}_i - \mathbf{t}_j\right|, \ \sigma_{\text{t},ij}\right) + \varepsilon_{\text{angle}} \text{Ł}\left(\left|\mathbf{b}_i - \mathbf{b}_j\right|, \ \sigma_{\text{b}}\right) +$$

$$\varepsilon_{\text{angle}} \text{Ł}\left(\left|\mathbf{b}_i - \mathbf{t}_j\right|, \ \sigma_{\text{tb}}\right) + \text{Ł}\left(\left|\mathbf{m}_i - \mathbf{m}_j\right|, \ \sigma_{\text{m}}\right) +$$

$$\sum_{\alpha,\beta}^{N_{\text{a}i}, N_{\text{a}j}} \varepsilon_{\text{hh}} \mathcal{M}\left(\left|\mathbf{a}_{i\alpha} - \mathbf{a}_{j\beta}\right|, \ r_0, \varrho, r_{\text{cut}}^{\text{att}}\right).$$

$$\tag{S26}$$

The function Ł is defined as the repulsive component of the Lennard-Jones potential shifted to zero at the interaction diameter:

$$\text{Ł}(x, \sigma) \equiv \theta(\sigma - x) \left[\left(\frac{\sigma}{x}\right)^{12} - 1\right] \tag{S27}$$

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

$$\mathcal{M}(x, r_0, \varrho, r_{\text{cut}}) = \theta(r_{\text{cut}} - x) \times$$

$$\left[\left(e^{\varrho\left(1 - \frac{x}{r_0}\right)} - 2\right) e^{\varrho\left(1 - \frac{x}{r_0}\right)} - V_{\text{shift}}(r_{\text{cut}})\right] \tag{S28}$$

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

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

**Subunit-subunit interaction parameter values.** *Attractors:* The strength of attractive interactions is parameterized by the well-depth $\varepsilon_{\text{hh}}$ for a pair of attractors on subunits as follows. Subunit-subunit edge attractor pairs have a well-depth of $\varepsilon_{\text{hh}}$. 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 $r_{cut}^{att} = 2.0$ 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,[16] 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 $\sigma_b$ and the 'Top' - 'Top' interaction $\sigma_t$ 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 $\alpha$-carboxysomes has not been experimentally estimated, but measurements on $\beta$-carboxysome shells obtained $\kappa \approx 25k_B T$,[17] which is smaller than for a typical viral capsid. As noted above, our simulated shell subunits, with diameter 13 nm, are larger than the hexamer size (7 nm diameter) in a carboxysome. To maintain the characteristic elastic length scale $R_E = \sqrt{2a\kappa}$ of the simulated shells approximately equal to that of $\beta$-carboxysome shells $R_E \approx 47$ nm, despite the large size of model subunits, we set $\varepsilon_{angle} = 0.5k_B T$, which corresponds to a bending modulus of about 5-10 $k_B T$,[3] and thus $R_E \approx 38 - 54$ nm.

**Scaffold-subunit interactions.** The scaffold-subunit interaction is modeled by a short-range repulsion between pairs of scaffold beads and scaffold-excluder psudoatoms on subunits representing the excluded volume, plus an attractive interaction between pairs of shell-interacting beads in scaffolds and subunit 'Bottom' pseudoatoms. For subunit $i$ with excluder positions $\{\mathbf{x}_{i\alpha}\}$ and 'Bottom' psuedoatom 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:

$$U_{\text{sh}}(\{\mathbf{x}_{i\alpha}\}, \mathbf{b}_i, \{\mathbf{s}_{j\beta}\}, \{\mathbf{s}_{j\gamma}\}) =$$

$$\sum_{\beta}^{N_{\text{s}}} \sum_{\alpha}^{N_{\text{x}}} \text{Ł}\left(|\mathbf{x}_{i\alpha} - \mathbf{s}_{j\beta}|, \sigma_{\text{exs}}\right) +$$

$$\sum_{\gamma}^{N_{\text{sh}}} \varepsilon_{\text{sh}} \mathcal{M}\left(\left|\mathbf{b}_i - \mathbf{s}_{j\gamma}\right|, r_0, \varrho^{\text{sh}}, r_{\text{cut}}^{\text{sh}}\right) \tag{S29}$$

where $\varepsilon_{\text{sh}}$ parameterizes the scaffold-subunit interaction strength, $N_{\text{s}}$ and $N_{\text{sh}}$ are the total number of beads and the number of shell-interacting beads in each scaffold, $N_{\text{x}}$ is the numbers of excluders on a shell subunit, $\sigma_{\text{exs}} = 0.375$ and $\sigma_{\text{t}} = 0.375$ 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

$$U_{\text{sc}}(\mathbf{c}_i, \{\mathbf{s}_{j\alpha}\}, \{\mathbf{s}_{j\beta}\}) =$$

$$\sum_{\alpha}^{N_{\text{s}}} \text{Ł}\left(\left|\mathbf{c}_i - \mathbf{s}_{j\alpha}\right|, \sigma_{\text{sc}}\right) + \tag{S30}$$

$$\sum_{\beta}^{N_{sc}} \varepsilon_{\text{sc}} \mathcal{L}\left(\left|\mathbf{c}_i - \mathbf{s}_{j\beta}\right|, \sigma_{\text{sc}}, r_{\text{cut}}^{\text{sc}}\right) \tag{S31}$$

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

$$\mathcal{L}(x, \sigma, r_{\text{cut}}) = \theta(x - r_{\text{cut}}) \times$$

$$\left\{4\left[\left(\frac{x}{\sigma}\right)^{1}2 - \left(\frac{x}{\sigma}\right)^{6}\right] - V_{\text{shift}}(r_{\text{cut}})\right\} \tag{S32}$$

with $\varepsilon_{\text{sc}}$ an adjustable parameter that sets the strength of the scaffold-cargo interaction, $N_{\text{s}}$ as the total number of scaffold beads, $N_{\text{sc}}$ 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

cutoff 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:

$$U_{ss}(\mathbf{R}_i, \mathbf{R}_j) =$$
$$\mathrm{K_{bond}}\left(\mathrm{R}_{ij}, \sigma_s, \kappa\right) : \{i, j\} \text{ bonded}$$
$$\varepsilon_{ss}\mathcal{L}\left(\mathrm{R}_{ij}, \sigma_{ss}, r_{\mathrm{cut}}^{ss}\right) : \{i, j\} \text{ nonbonded} \tag{S33}$$

where $R_{ij} \equiv |\mathbf{R}_i - \mathbf{R}_j|$ is the center-to-center distance between the scaffold beads, $\varepsilon_{ss}$ is the well-depth of the Lennard-Jones (Eq. S32) potential, $\sigma_{ss} = 0.5$ is the effective scaffold-scaffold excluded volume size, and $r_{\mathrm{cut}}^{ss} = 3\sigma_{ss}$ is the interaction cutoff length.

Bonds are represented by a harmonic potential:

$$\mathrm{K_{bond}}(\mathrm{R}_{ij}, \sigma, \kappa) \equiv \frac{k_{\mathrm{bond}}}{2}(\mathrm{R}_{ij} - \sigma)^2. \tag{S34}$$

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

$$U_{cc}(\mathbf{c}_i, \mathbf{c}_j) = \mathbb{E}\left(|\mathbf{c}_i - \mathbf{c}_j|, \sigma_c\right). \tag{S35}$$

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

$$U_{hc}(\{\mathbf{x}_{i\alpha}\}, \mathbf{c}_j) = \sum_{\alpha}^{N_x} \mathbb{E}\left(|\mathbf{x}_{i\alpha} - \mathbf{c}_j|, \sigma_{\mathrm{exc}}\right). \tag{S36}$$
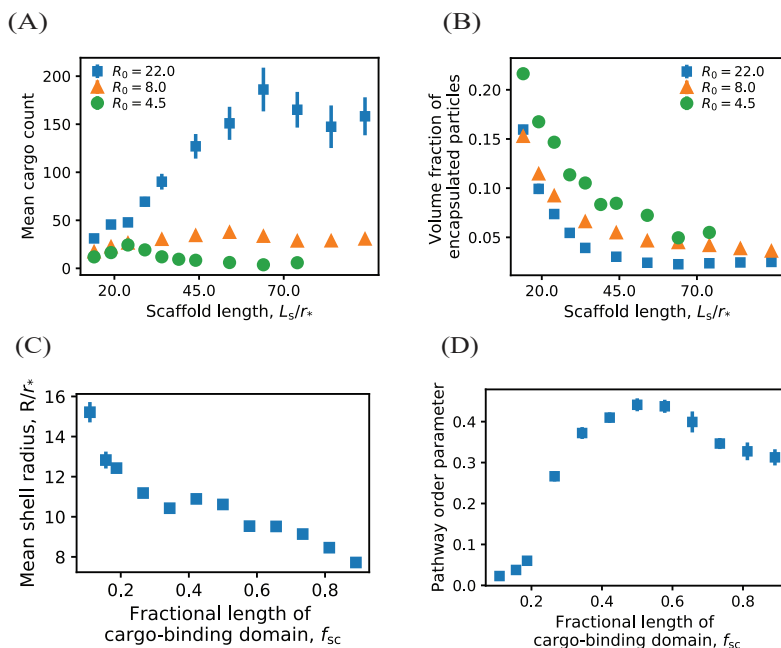
Figure S3: **(A)** Dependence of the number of encapsulated cargo particles on the scaffold length for the parameter sets considered in Fig. 3 of the main text. **(B)** Dependence of the volume fraction of encapsulated particles (cargo and scaffolds) on the scaffold length for the same parameters as in (A). **(C)** Dependence of the shell size on the the scaffold cargo-binding domain fraction $f_{sc} = L_{sc}/L_s$, for $R_0 = 22$, $L_s = 64$, and $L_{sh} = 7$. **(D)** The pathway order parameter, defined as the maximum number of unassembled shell subunits adsorbed on scaffold-cargo aggregates at any point during a trajectory, normalized by the size of the final shell, is shown as a function of the cargo-binding domain fraction $f_{sc}$. The pathway transitions to two-step assembly at $f_{sc} \approx 0.3$. The order parameter decreases slightly at high values of $f_{sc}$ because the large scaffold-cargo valence leads to rapid nucleation of multiple small cargo-scaffold complexes, leading to fewer adsorbed shell subunits per complex. Other parameter values for (A-D) are as follows. Shell subunit-subunit affinities: $\varepsilon_{hh} = 3.15$ at $R_0 = 4.5$, $\varepsilon_{hh} = 2.85$ at $R_0 = 8.0$ and $\varepsilon_{hh} = 2.65$ at $R_0 = 22$; scaffold-subunit affinity $\varepsilon_{sh} = 2.5$, and scaffold-cargo affinity $\varepsilon_{sc} = 1.0$.
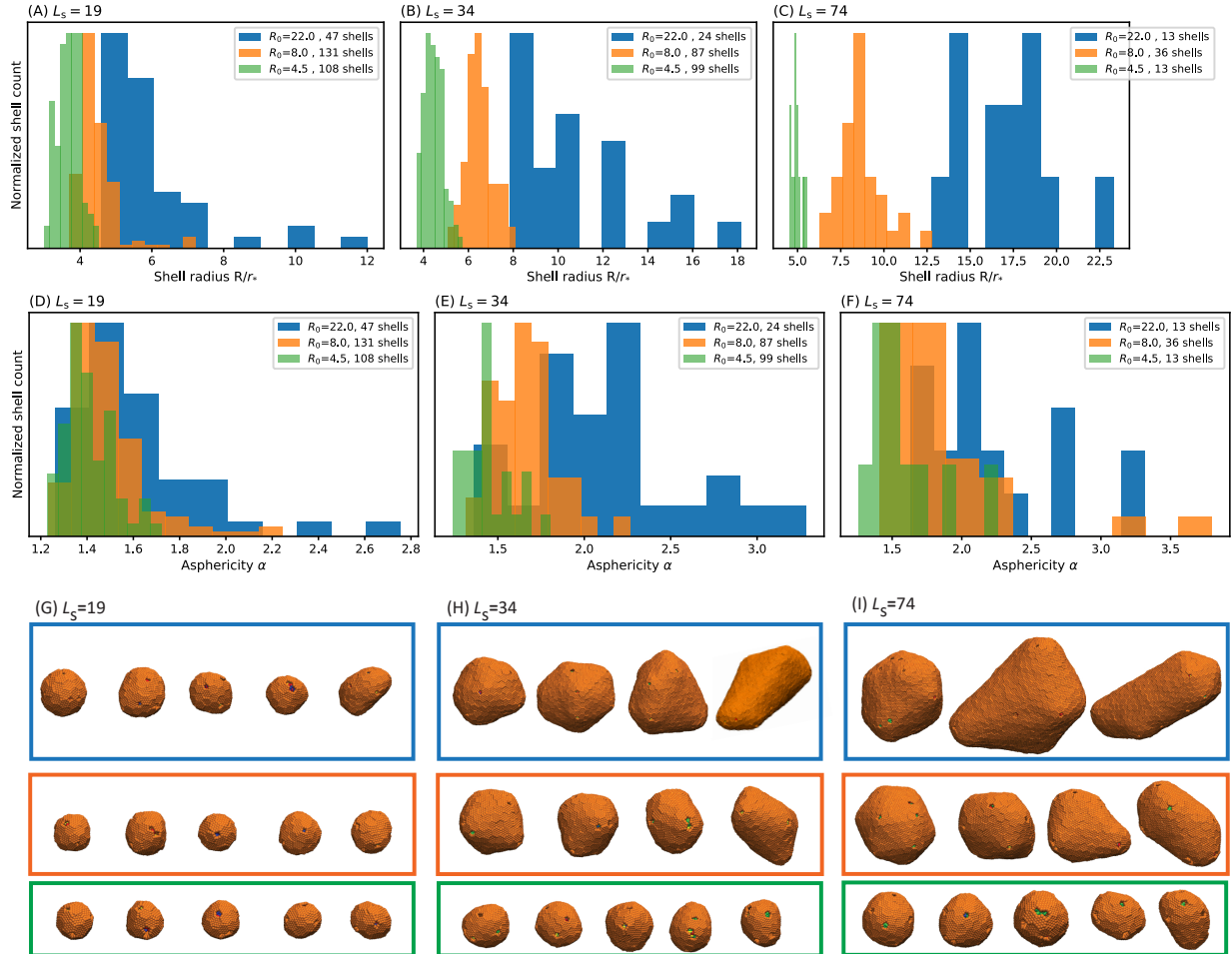
15

Figure S4: **(A-C)** Distribution of shell sizes at three scaffold lengths **(A)** $L_s = 19$, **(B)** $L_s = 34$, and **(C)** $L_s = 74$. Each plot shows distributions for the three shell spontaneous curvature radii considered in Fig. 3 of the main text. **(D-F)** Distribution of asphericity of shells at the same parameter sets. Asphericity is defined as the $\alpha = R_{max}/R_{min}$ with $R_{max}$ and $R_{min}$ the maximum and minimum radius of shell subunits measured from the center of mass of the shell. **(G-I)** Snapshots of assembly products for each of the three scaffold lengths. The blue (top), orange (middle), and green (bottom) boxes correspond to spontaneous curvature radii of $R_0 = 22.0$, $8.0$, and $4.5$ respectively.
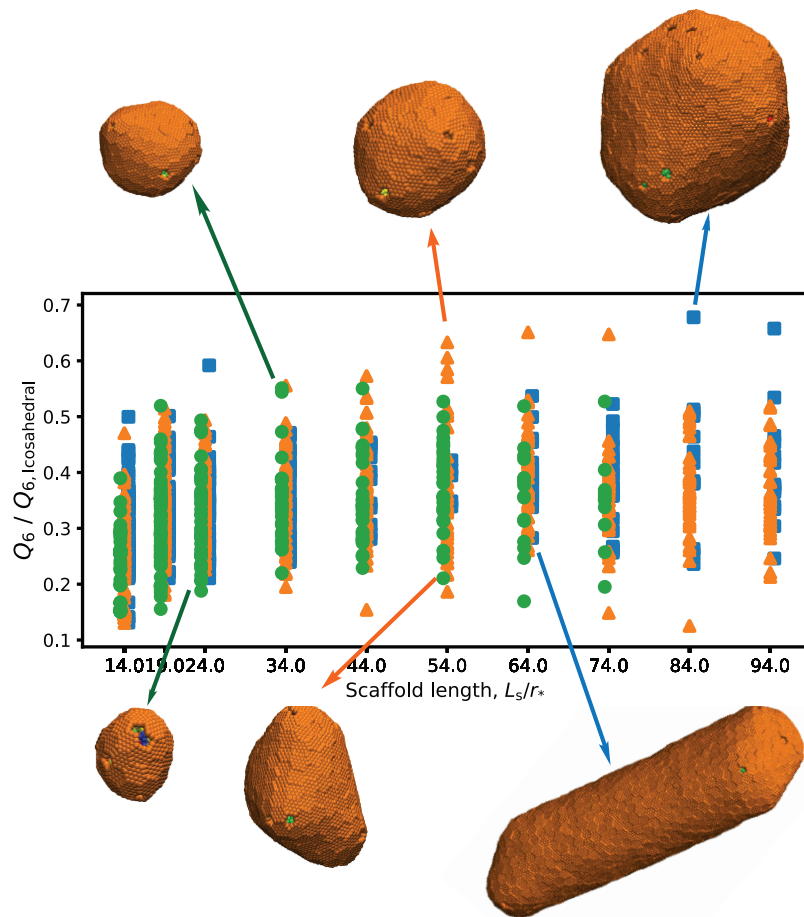
16

Figure S5: The icosahedral bond order parameter Q6 of Ref.[18] as a function of scaffold length for $R_0 = 22.0$, 8.0 and 4.5. Q6 is normalized by the value for perfect icosahedral symmetry, Q6 = 0.663. Snapshots of typical shells with indicated Q6 values are shown. Elongated shells, conical shells, and small shells with double vacancies have low values of icosahedral order.
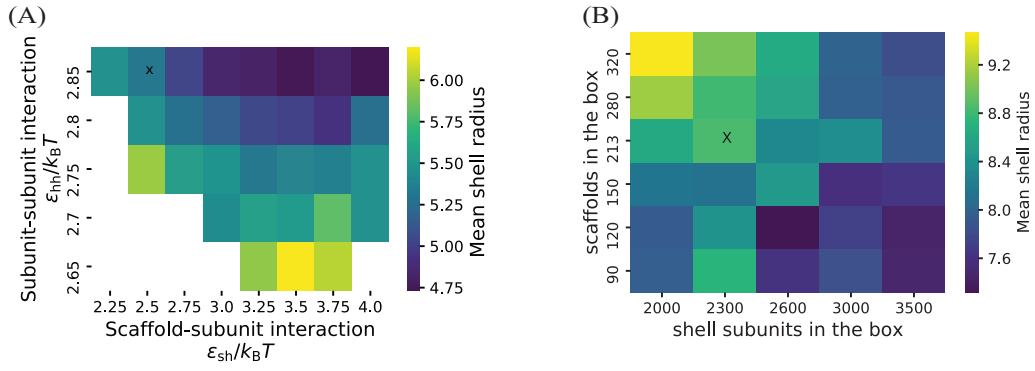
Figure S6: **(A)** Dependence of the mean shell radius on the scaffold-subunit and subunit-subunit affinities. The scaffold length is $L_s = 24$ ($L_{sc} = 7$, $L_{sm} = 10$), the shell spontaneous curvature radius is $R_0 = 8.0$ and the scaffold-cargo affinity is $\varepsilon_{sc} = 1 k_B T$. Note that in the large $L_s$ limit where relatively few scaffold molecules are packaged, strong subunit-subunit interactions ($\varepsilon_{ss} \gtrsim 3.15$ at $R_0 = 4.5$, $\varepsilon_{ss} \gtrsim 2.85$ at $R_0 = 8$ and $\varepsilon_{ss} \gtrsim 2.65$ at $R_0 = 22$) are required for complete shell assembly. **(A)** 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 $R_0 = 8.0$, and the scaffold-cargo affinity is $\varepsilon_{sc} = 1.0 k_B T$. 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.
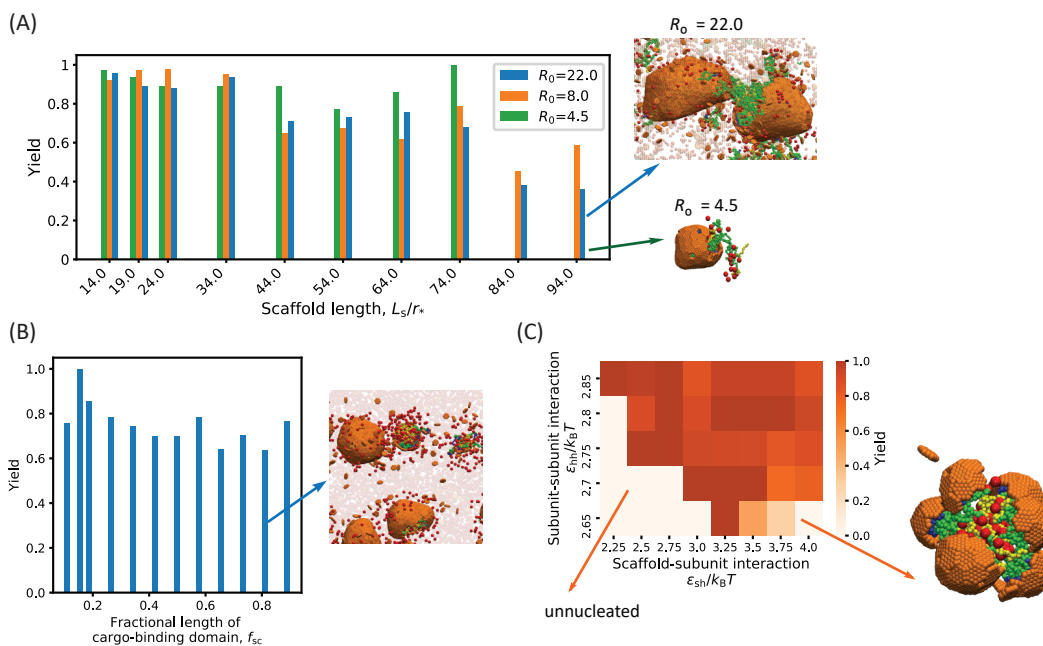
Figure S7: **(A)** Fraction of subunits in complete shells for the simulations corresponding to Fig. 3A,B (main text). We categorize shells as 'complete' according to the following criteria. A complete icosahedral shell composed of $n_{hex}$ hexamers and 12 pentamers has $6n_{hex} + 30$ bonds, with bonds defined as strong interactions between neighboring subunits. However, since our simulations do not include pentamers, a complete shell has 12 five-fold vacancies, and thus $6n_{hex} - 30$ bonds. In the case of small shells, we also observed assembly of two classes of complete, but non-icosahedral shells: shells that are missing two adjacent pentamers, and shells with tetrameric holes. Finally, very large shells require extremely long simulation times to assemble completely. Therefore, we consider shells with more than 300 subunits to be complete if they are missing up to 3 subunits. An example of an incomplete shell at $L_s = 94$ and $R_0 = 22.0$ is shown next to the plot. In this case shell completion did not occur on simulation timescales because multiple nuclei formed and the concentration of free subunits was depleted before assembly completed. The lower snapshot to the right of the plot shows an example of a typical failure mode for scaffold molecules that are much longer than the spontaneous curvature radius ($L_s = 94$ and $R_0 = 4.5$). **(B)** Fraction of subunits in complete shells for the simulations corresponding to Fig. 3C (main text). The snapshot to the right of the plot shows an example of multi-nucleation at $f_{sc} = 0.81$, $L_s = 64$ and $R_0 = 22$. **(C)** Fraction of subunits in complete shells for the simulations corresponding to Fig. S6, shown as a function of the subunit-subunit and scaffold-subunit binding affinities. Because we have focused on relatively weak subunit-subunit affinities, we observe high yields for all values of $\varepsilon_{hh}$ in this range. However, combining high scaffold-subunit interactions with low subunit-subunit interactions results in a kinetic trap corresponding to multiple incomplete shells attached to a cluster of scaffolds and cargo. The snapshot to the right of the plot shows an example of this behavior for $\varepsilon_{hh} = 2.65k_BT$ and $\varepsilon_{sh} = 4.0k_BT$.

19

# References

1. Hagan, M. F.; Grason, G. M. Equilibrium Mechanisms of Self-Limiting Assembly. *arXiv.2007.01927* **2020**, https://arxiv.org/abs/2007.01927.

2. Perlmutter, J. D.; Mohajerani, F.; Hagan, M. F. Many-Molecule Encapsulation by an Icosahedral Shell. *eLife* **2016**, *5*, e14078.

3. Mohajerani, F.; Hagan, M. F. The Role of the Encapsulated Cargo in Microcompartment Assembly. *PLoS Comput. Biol.* **2018**, *14*, e1006351.

4. Lidmar, J.; Mirny, L.; Nelson, D. R. Virus Shapes and Buckling Transitions in Spherical Shells. **2003**, *68*, 051910.

5. Nguyen, T. T.; Bruinsma, R. F.; Gelbart, W. M. Elasticity Theory and Shape Transitions of Viral Shells. *Phys. Rev. E* **2005**, *72*, 051923.

6. Li, S.; Roy, P.; Travesset, A.; Zandi, R. Why Large Icosahedral Viruses Need Scaffolding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10971–10976.

7. Li, S.; Zandi, R.; Travesset, A.; Grason, G. M. Ground States of Crystalline Caps: Generalized Jellium on Curved Space. *Phys. Rev. Lett.* **2019**, *123*, 145501.

8. Schwartz, R.; Shor, P. W.; Prevelige, P. E.; Berger, B. Local Rules Simulation of the Kinetics of Virus Capsid Self-Assembly. *Biophys. J.* **1998**, *75*, 2626–2636.

9. Hagan, M. F.; Chandler, D. Dynamic Pathways for Viral Capsid Assembly. *Biophys. J.* **2006**, *91*, 42–54.

10. Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. Reversible Self-Assembly of Patchy Particles into Monodisperse Icosahedral Clusters. *J. Chem. Phys.* **2007**, *127*, 085106.

11. Baschek, J. E.; Klein, H. C. R.; Schwarz, U. S. Stochastic Dynamics of Virus Capsid Formation: Direct *versus* Hierarchical Self-Assembly. *BMC biophysics* **2012**, *5*, 1–18.

12. Nguyen, H. D.; Reddy, V. S.; Brooks, C. L. Deciphering the Kinetic Mechanism of Spontaneous Self-Assembly of Icosahedral Capsids. *Nano Lett.* **2007**, *7*, 338–344.

13. Rapaport, D. Self-Assembly of Polyhedral Shells: A Molecular Dynamics Study. *Phys. Rev. E.* **2004**, *70*, 051905.

14. Elrad, O.; Hagan, M. F. Encapsulation of a Polymer by an Icosahedral Virus. *Phys. Biol.* **2010**, *7*, 045003.

15. Hagan, M. F. Modeling Viral Capsid Assembly. *Adv. Chem. Phys.* **2014**, *155*, 1–68.

16. Johnston, I. G.; Louis, A. A.; Doye, J. P. K. Modelling the Self-Assembly of Virus Capsids. *J. Phys.: Condens. Matter* **2010**, *22*, 104101.

17. Faulkner, M.; Rodriguez-Ramos, J.; Dykes, G. F.; Owen, S. V.; Casella, S.; Simpson, D. M.; Beynon, R. J.; Liu, L.-N. Direct Characterization of the Native Structure and Mechanics of Cyanobacterial Carboxysomes. *Nanoscale* **2017**, *9*, 10662–10673.

18. Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. Bond-Orientational Order in Liquids and Glasses. *Physical Review B* **1983**, *28*, 784.