RESEARCH ARTICLE

Many-molecule encapsulation by an icosahedral shell

Jason D Perlmutter[†], Farzaneh Mohajerani[†], Michael F Hagan*

Martin Fisher School of Physics, Brandeis University, Waltham, United States

Abstract We computationally study how an icosahedral shell assembles around hundreds of molecules. Such a process occurs during the formation of the carboxysome, a bacterial microcompartment that assembles around many copies of the enzymes ribulose 1,5-bisphosphate carboxylase/ oxygenase and carbonic anhydrase to facilitate carbon fixation in cyanobacteria. Our simulations identify two classes of assembly pathways leading to encapsulation of many-molecule cargoes. In one, shell assembly proceeds concomitantly with cargo condensation. In the other, the cargo first forms a dense globule; then, shell proteins assemble around and bud from the condensed cargo complex. Although the model is simplified, the simulations predict intermediates and closure mechanisms not accessible in experiments, and show how assembly can be tuned between these two pathways by modulating protein interactions. In addition to elucidating assembly pathways and critical control parameters for microcompartment assembly, our results may guide the reengineering of viruses as nanoreactors that self-assemble around their reactants.
DOI: 10.7554/eLife.14078.001

*For correspondence: hagan@brandeis.edu

[†]These authors contributed equally to this work

Introduction

Encapsulation is a hallmark of biology. A cell must co-localize high concentrations of enzymes and reactants to perform the reactions that sustain life, and it must safely store genetic material to ensure long-term viability. While lipid-based organelles primarily fulfill these functions in eukaryotes, self-assembling protein shells take the lead in simpler organisms. For example, viruses surround their genomes with a protein capsid, while bacteria use large icosahedral shells known as bacterial micro-compartments (BMCs) to sequester the enzymes and reactions responsible for particular metabolic pathways (*Kerfeld et al., 2010*; *Axen et al., 2014*; *Shively et al., 1998*; *Bobik et al., 1999*; *Erbilgin et al., 2014*; *Petit et al., 2013*; *Price and Badger, 1991*; *Shively et al., 1973*; *Shively et al., 1973*; *Kerfeld and Erbilgin, 2015*). Within diverse bacteria, BMC functions have been linked to bacterial growth, carbon fixation, symbiosis, or pathogenesis (*Kerfeld and Erbilgin, 2015*). Other protein-based compartments are found in bacteria and archea (*e.g.* encapsulins (*Sutter et al., 2008*) and gas vesicles (*Pfeifer, 2012*; *Sutter et al., 2008*)) and even eukaryotes (*e.g.* vault particles (*Kickhoefer et al., 1998*)), while some viruses may assemble around lipidic globules (*Lindenbach and Rice, 2013*; *Faustino et al., 2014*). Thus, understanding the factors that control microcompartment assembly and encapsulation is a central question in modern cell biology. From the perspectives of synthetic biology and nanoscience, there is great interest in reengineering BMCs or viruses as nanoreactors that spontaneously encapsulate enzymes and reagents in vitro (*e.g.* *Luque et al., 2014*; *Douglas and Young, 1998*; *Rurup et al., 2014*; *Patterson et al., 2014*; *Patterson et al., 2012*; *Zhu et al., 2014*; *Rhee et al., 2011*; *Rurup et al., 2014*; *Wörsdörfer et al., 2012*; *Comas-Garcia et al., 2014*), or as customizable organelles that assemble around a program-mable set of core enzymes in vivo, introducing capabilities such as carbon fixation or biofuel produc-tion into bacteria or other organisms (*e.g.* *Kerfeld and Erbilgin, 2015*; *Bonacci et al., 2012*; *Parsons et al., 2010*; *Choudhary et al., 2012*; *Lassila et al., 2014*). However, the principles

eLife digest Bacterial microcompartments are protein shells that are found inside bacteria and enclose enzymes and other chemicals required for certain biological reactions. For example, the carboxysome is a type of microcompartment that enables the bacteria to convert the products of photosynthesis into sugars. During the formation of a microcompartment, the outer protein shell assembles around hundreds of enzymes and chemicals. This formation process is tightly controlled and involves multiple interactions between the shell proteins and the cargo – the enzymes and other reaction ingredients – they will enclose. Understanding how to control which enzymes are encapsulated within microcompartments could help researchers to re-engineer the microcompartments so that they contain drugs or other useful products.

Recent studies have used microscopy to visualize how microcompartments are assembled. However, most of the intermediate structures that form during assembly are too small and short-lived to be seen. It has therefore not been possible to explore in detail how shell proteins collect the necessary cargo and then assemble into an ordered shell with the cargo on the inside. Experiments alone are probably not enough to understand the process, especially since microcompartment assembly can currently only be studied within live cells or cellular extract. Within these complex environments it is difficult to determine the effect of any individual factor on the overall assembly process.

Perlmutter, Mohajerani and Hagan have now taken a different approach by developing computational and theoretical models to explore how microcompartments assemble. Computer simulations showed that microcompartments could assemble by two pathways. In one pathway, the protein shell and cargo coalesce at the same time. In the other pathway, the cargo molecules first assemble into a large disordered complex, with the shell proteins attached on the outside. The shell proteins then assemble, carving out a piece of the cargo complex. The simulations showed that many factors affect how the shell assembles, such as the strengths of the interactions between the shell proteins and the cargo. They also identified a factor that controls how much cargo ends up inside the assembled shell.

Perlmutter, Mohajerani and Hagan found that, in addition to revealing how microcompartments may assemble within their natural setting, the simulations provided guidance on how to re-engineer microcompartments to assemble around other components. This would enable researchers to create customizable compartments that self-assemble within bacteria or other host organisms, for example to carry out carbon fixation or make biofuels.

A future challenge will be to investigate other aspects of microcompartment assembly, such as the factors that control the size of these compartments.

controlling such co-assembly processes have yet to be established, and it is not clear how to design systems to maximize encapsulation.

In this article we take a step toward this goal, by developing theoretical and computational models that describe the dynamical encapsulation of hundreds of cargo molecules by self-assembling icosahedral shells. Although our models are general, we are motivated by recent experiments on a type of BMC known as the carboxysome (*Kerfeld et al., 2010*; *Schmid et al., 2006*; *Iancu et al., 2007*; *Tanaka et al., 2008*). Carboxysomes are large (40–400 nm), roughly icosahedral shells that encapsulate a dense complex of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and other proteins to facilitate the Calvin-Bensen-Bassham cycle in autotrophic bacteria (*Price and Badger, 1991*; *Shively et al., 1973*; *Shively et al., 1973*; *Iancu et al., 2007*; *2010*; *Kerfeld et al., 2010*; *Tanaka et al., 2008*). Recently, striking microscopy experiments visualized $\beta-$carboxysome shells assembling on and budding from procarboxysomes (the condensed complex of RuBisCO and other proteins found in the interior of carboxysomes) (*Cameron et al., 2013*; *Chen et al., 2013*). Genomic analysis suggests that many BMCs with diverse functions assemble via similar pathways (*Cameron et al., 2013*; *Kerfeld and Erbilgin, 2015*). However, the mechanisms of budding and pinch-off to close the shell remain incompletely understood because of the small size and transient nature of assembly intermediates. Moreover, experiments suggest that
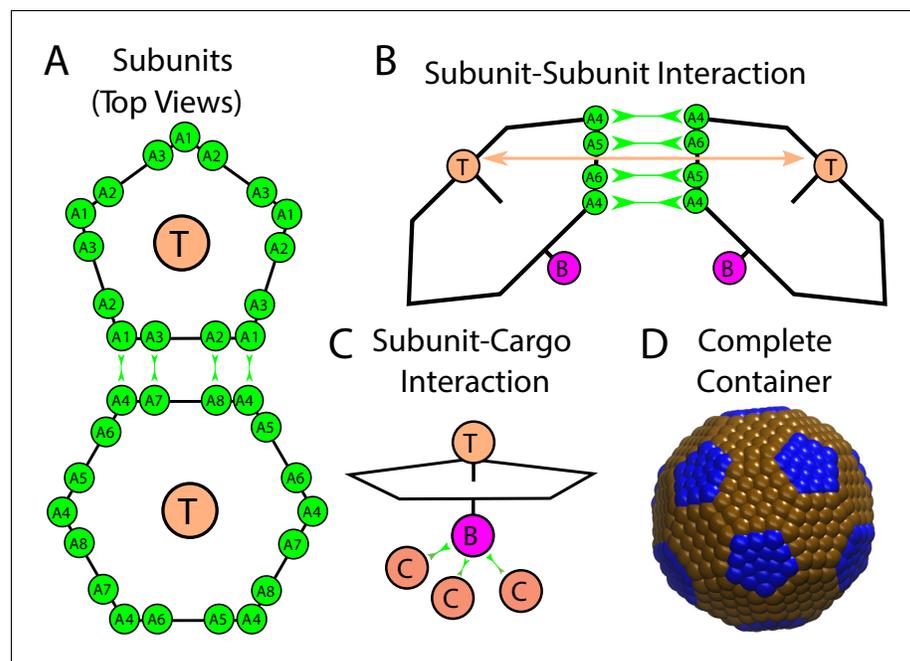
Figure 1. Description of the model. (**A**) Each shell subunit contains 'Attractors' (green circles) on the perimeter, a 'Top' (tan circle, 'T') in the center above the plane, and a 'Bottom' (purple circle, 'B' below the plane). (**B**) Interactions between complementary Attractors drive subunit dimerization, with the Top-Top repulsions (tan arrow) tuned to favor the subunit-subunit angle in a complete shell. Complementary pairs of attractors are indicated by green arrows in (**A**) for the pentamer-hexamer interface and in (**B**) for the hexamer-hexamer interface. (**C**) Bottom psuedoatoms bind cargo molecules (terra cotta circles, '**C**'), while excluder atoms (blue and brown pseudoatoms in (**D**)) placed in the plane of the pentagon experience excluded volume interactions with the cargo. (**D**) The positions of excluder atoms in the lowest energy shell geometry, a truncated icosahedron with 12 pentamers (blue) and 20 hexamers (brown).

DOI: 10.7554/eLife.14078.003

$\alpha-$carboxysomes (another form of carboxysome) assemble by a different mechanism, in which shell assembly encapsulates an initially diffuse pool of RuBisCO (*Iancu et al., 2010*; *Cai et al., 2015*). The factors determining which of these assembly pathways occurs are unknown.

BMC assembly is driven by a complex interplay of interactions among the proteins forming the external shell and the interior cargo. It is difficult, with experiments alone, to parse these interactions for those mechanisms and factors that critically influence assembly pathways, especially due to the lack of an in vitro assembly system. Models which can correlate individual factors to their effect on assembly are therefore an important complement to experiments.

Previous experimental and theoretical studies of encapsulation by icosahedral shells, e.g. the assembly of viral capsids around their nucleic acid genomes (*e.g. Hu and Shklovskii, 2007*; *Kivenson and Hagan, 2010*; *Elrad and Hagan, 2010*; *Perlmutter et al., 2013*; *2014*; *Mahalik and Muthukumar, 2012*; *Zhang et al., 2013*; *Zhang and Linse, 2013*; *Hagan, 2008*; *Devkota et al., 2009*; *Dixit et al., 2006*; *Borodavka et al., 2012*; *Dykeman et al., 2013*; *2014*; *Zlotnick et al., 2013*; *Johnson et al., 2004*; *Patel et al., 2015*; *Cadena-Nava et al., 2012*; *Comas-Garcia et al., 2012*; *2014*; *Garmann et al., 2014a*; *2014b*; *Malyutin and Dragnea, 2013*), have demonstrated that the structure of the cargo can strongly influence assembly pathways and products. However, BMCs assemble around a cargo which is topologically different from a nucleic acid — a fluid complex comprising many, noncovalently linked molecules. We demonstrate here that changing the cargo topology leads to new assembly pathways and different critical control parameters.

We present phase diagrams and analysis of dynamical simulation trajectories showing how the thermodynamics, assembly pathways, and emergent structures depend on the interactions among shell proteins and cargo molecules. Within distinct parameter ranges, we observe two classes of assembly pathways, which resemble those suggested for respectively $\alpha-$ or $\beta-$carboxysomes. We
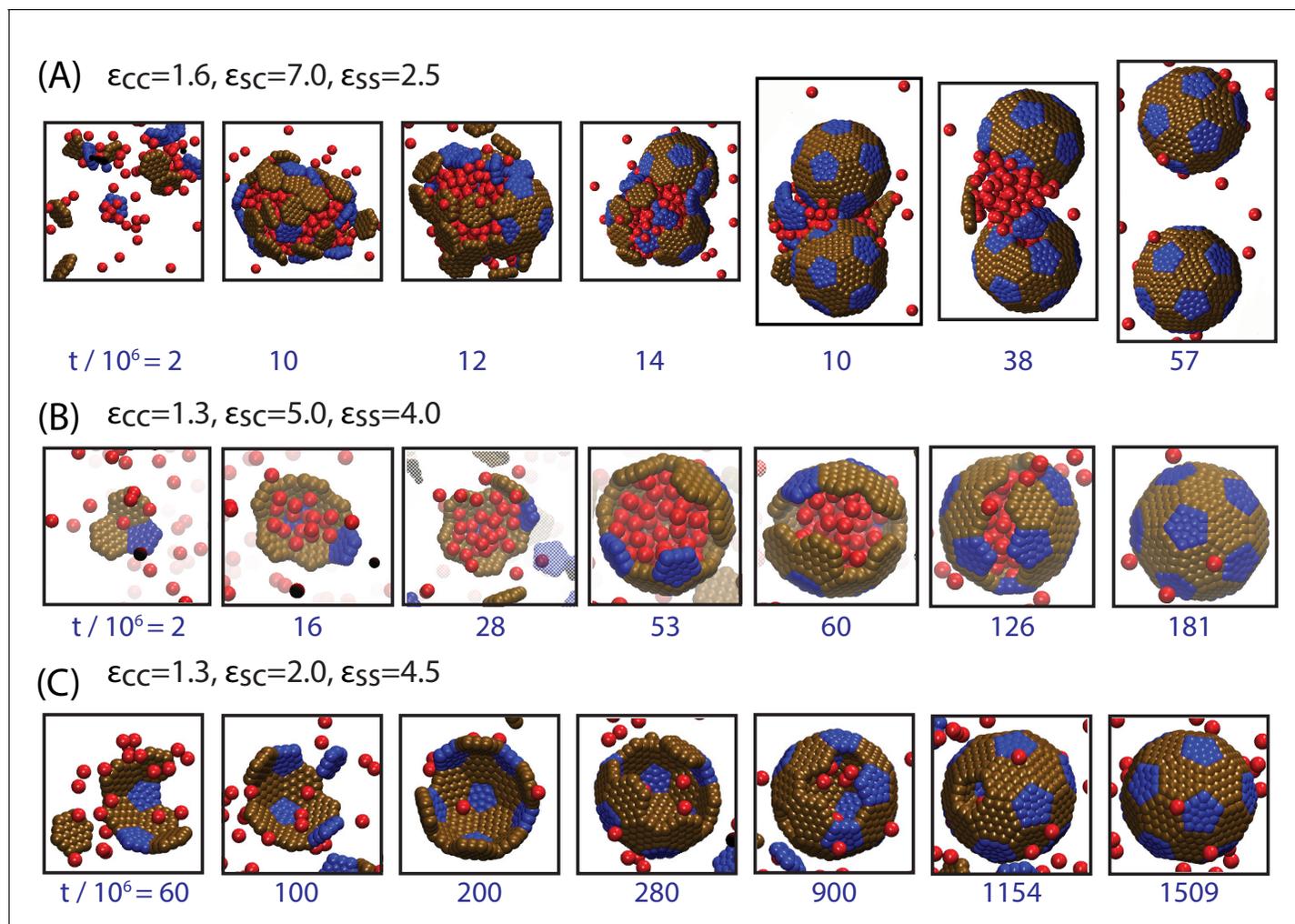
Figure 2. Snapshots illustrating typical assembly trajectories. (A) Multi-step assembly involving an amorphous globule of cargo and shell subunits. (B) Single-step assembly, in which shell assembly drives local cargo condensation. and (C) when shell-cargo interactions are too weak to condense the cargo. The values of the cargo-cargo (ε_{CC}), shell subunit-cargo (ε_{SC}), and subunit-subunit (ε_{SS}) interaction strengths are listed above each panel (all energies are in units of the thermal energy k_BT), and the time (in units of 10^6 timesteps) is noted below each image. The color scheme here and throughout the manuscript is: Red=Cargo, Blue=Pentagon Excluder, Brown=Hexagon Excluder. Attractor and Bottom pseudoatoms are omitted to aid visibility. Videos of assembly trajectories are included below.

DOI: 10.7554/eLife.14078.004

The following figure supplements are available for figure 2:

Figure supplement 1. Snapshots from additional trajectories, including a trajectory with a pre-equilibrated cargo globule.
DOI: 10.7554/eLife.14078.005

Figure supplement 2. Snapshots from assembly trajectories around anisotropic cargo particles, for (A) strong cargo-cargo interactions leading to two-step, globule-mediated assembly, and (B) weak cargo-cargo interactions leading to simultaneous assembly and cargo condensation.
DOI: 10.7554/eLife.14078.006

find that tunability of cargo loading is a key functional difference between the two classes of pathways. Shells assembled around a diffuse cargo can be varied from empty (containing almost no cargo) to completely full, whereas assembly around a condensed, procarboxysome-like complex invariably produces full shells. While we find that the encapsulated cargo becomes ordered due to confinement, complete crystalline order in the globule before encapsulation inhibits budding. We discuss these results in the context of recent observations on carboxysome assembly, and their implications for engineering BMCs, viruses or drug delivery vehicles that assemble around a fluid cargo (*e.g.* Refs. [*Kerfeld and Erbilgin, 2015*; *Parsons et al., 2010*; *Choudhary et al., 2012*; *Lassila et al., 2014*; *Luque et al., 2014*; *Douglas and Young, 1998*; *Rurup et al., 2014*; *Patterson et al., 2014*;
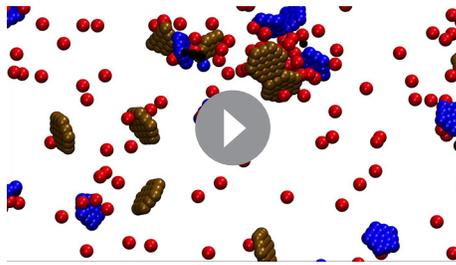
Video 1. Animation of a typical simulation showing assembly around a cargo globule. Parameters are $\varepsilon_{CC} = 1.6$, $\varepsilon_{SC} = 7$, and $\varepsilon_{SS} = 2.5$.
DOI: 10.7554/eLife.14078.007

Patterson et al., 2012; *Zhu et al., 2014*; *Rhee et al., 2011*; *Rurup et al., 2014*; *Wörsdörfer et al., 2012*]).
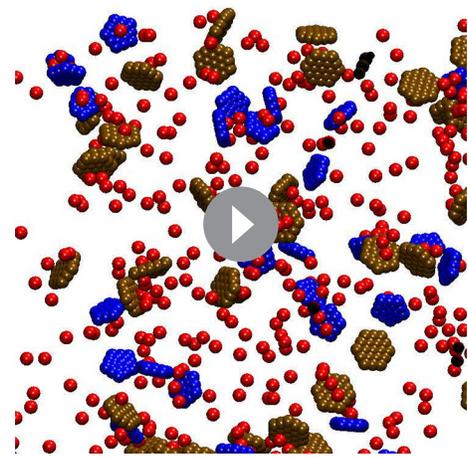
Video 2. Animation of a typical simulation showing simultaneous assembly and cargo condensation. Parameters are $\varepsilon_{CC} = 1.3$, $\varepsilon_{SC} = 9$, and $\varepsilon_{SS} = 3.5$.
DOI: 10.7554/eLife.14078.008

Results

Our model system is motivated by icosahedral viral capsids and BMCs (*Tanaka et al., 2008*; *Kerfeld et al., 2010*). Since icosahedral symmetry can accommodate at most 60 identical subunits, formation of large icosahedral structures requires subunits to assemble into different local environments. The subunits can be grouped into pentamers and hexamers, with 12 pentamers at the icosahedron vertices and the remaining subunits in hexamers. Viruses typically assemble from small oligomers of the capsid protein, which we refer to as the basic assembly unit (*Hagan, 2014*). Recent AFM experiments demonstrated that hexamers are the basic assembly unit during the assembly of BMC shell facets (*Sutter et al., 2016*), and the carboxysome major shell proteins crystallize as pentamers and hexamers (*Tanaka et al., 2008*). Motivated by these observations, our model considers two basic assembly units, one a pentamer and the other a hexamer, with interactions designed so that the lowest energy structure corresponds to a truncated icosahedron with 12 pentamers and 20 hexamers (*Figure 1*). While BMCs generally have more hexamers, our model is intended to explore the general principles of assembly around a fluid cargo rather than model a specific system. Further details of the model and a thermodynamic analysis are given in section 3 and the appendices.

To understand how assembly around multiple cargo molecules depends on the relative strengths of interactions between components, we performed dynamical simulations as a function of the parameters controlling shell subunit-subunit (ε_{SS}), shell subunit-cargo (ε_{SC}), and cargo-cargo (ε_{CC}) interaction strengths. All energy values are given in units of the thermal energy, $k_B T$. We focus on parameters for which shell subunit-subunit interactions are too weak to drive assembly in the absence of cargo ($\varepsilon_{SS} \leq 4.5$). Except where mentioned otherwise, the cargo diameter is set equal to the circumradius of a shell subunit.

For the simulated density of cargo particles, the phase behavior (in the absence of shells) corresponds to a vapor at $\varepsilon_{CC} = 1.3$, liquid-vapor phase coexistence for $\varepsilon_{CC} \in [1.6, 2.0]$ (the phase coexistence boundary is slightly below $\varepsilon_{CC} = 1.6$), and a solid phase at $\varepsilon_{CC} = 3.0$. We find that tuning ε_{CC} through phase coexistence dramatically alters the typical assembly process. Strong cargo interactions ($\varepsilon_{CC} \geq 1.6$) drive formation of a globule followed by assembly and budding of a shell, such as observed for $\beta-$carboxysomes (*Figure 2A*, Simulation *Video 1*), while under weak interactions ($\varepsilon_{CC} < 1.6$) shell assembly usually proceeds in concert with cargo encapsulation (*Figure 2B*, Simulation *Video 2*), as suggested for assembly of $\alpha-$carboxysomes. We now elaborate on these classes of assembly pathways, and how the resulting assembly products depend on parameter values.

Assembly and budding from a cargo globule

We begin by discussing assembly behavior when the cargo-cargo interactions are strong enough to drive equilibrium phase coexistence ($\varepsilon_{CC} \geq 1.6$). Near the phase boundary ($\varepsilon_{CC} = 1.6$) a system of pure cargo particles is metastable on the timescales we simulate. However, for $\varepsilon_{SC} > 4$, adding shell
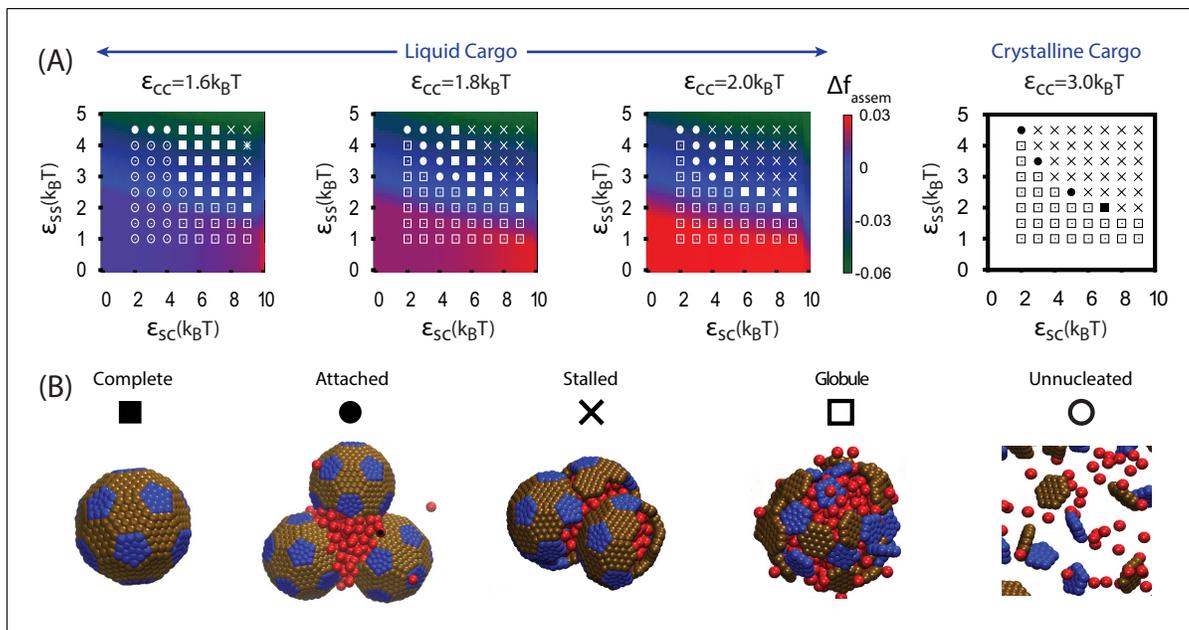
Figure 3. Results of assembly around a cargo globule. (**A**) The most frequently observed assembly outcome is overlaid on a color map of the theoretical free energy density difference $\Delta f_{\mathrm{assem}}$ (**Equation (3)**) between assembled shells and the unassembled globule. Results are plotted against the shell-cargo adsorption strength $\varepsilon_{\mathrm{SC}}$ and the shell-shell interaction strength $\varepsilon_{\mathrm{SS}}$ for indicated values of the cargo-cargo interaction strength $\varepsilon_{\mathrm{CC}}$. (**B**) Representative snapshots of the predominant assembly outcomes shown in (**A**).
DOI: 10.7554/eLife.14078.009

The following source data and figure supplements are available for figure 3:

Source data 1. List of all simulation outcomes for *Figures 3A,5A*.
DOI: 10.7554/eLife.14078.010

Source data 2. Criteria used to categorize assembly outcomes.
DOI: 10.7554/eLife.14078.011

Figure supplement 1. The distribution of assembly outcomes in *Figure 3A* is shown as a function of $\varepsilon_{\mathrm{SC}}$ for indicated values of $\varepsilon_{\mathrm{CC}}$ and $\varepsilon_{\mathrm{SS}}$.
DOI: 10.7554/eLife.14078.012

Figure supplement 2. Results of assembly around a pre-equilibrated cargo globule.
DOI: 10.7554/eLife.14078.013

Figure supplement 2—source data 1. List of all simulation outcomes for *Figure 3—figure supplement 1—2*.
~~Figure~~ supplement 3. The number of cargo particles packaged as a function of parameters.
DOI: 10.7554/eLife.14078.015

subunits drives nucleation of a cargo globule with shell subunits adsorbed on the surface. The subsequent fate of the globule depends on parameter values; typical simulation end-states are shown as a function of parameter values in *Figure 3*. For moderate interaction strengths ($2.5 \leq \varepsilon_{\mathrm{SS}} \leq 3.5$) the globule grows to a large size, typically containing at least twice the cargo molecules that can be packaged within a complete shell. Adsorbed shell subunits then reversibly associate to form ordered clusters. Once a cluster acquires enough inter-subunit interactions to be a stable nucleus, it grows by coagulation of additional subunits or other adsorbed clusters. For the parameter set corresponding to *Figure 2A*, nucleation is fast in comparison to cluster growth, and thus two nuclei grow simultaneously. The last three images show the system immediately preceding and following detachment of the lower shell. Missing only one of its 32 subunits, the shell is connected to the remainder of the droplet only by a narrow neck of cargo. Insertion of the final subunit breaks the neck and completes shell detachment. The complete shell contains 120–130 cargo particles, which is slighty above random close packing (≈ 120 particles) but below fcc density (≈ 150 particles, see appendix 1.2).

Increasing the shell-shell interaction strength drives faster shell assembly and closure, thus limiting the size of the globule before budding. For the largest interaction strength we simulated ($\varepsilon_{\mathrm{SS}} = 4.5$) the globule typically does not exceed the size of a single shell, and multiple globules nucleate within the simulation box (*Figure 2—figure supplement 1*). This observation could place an upper bound
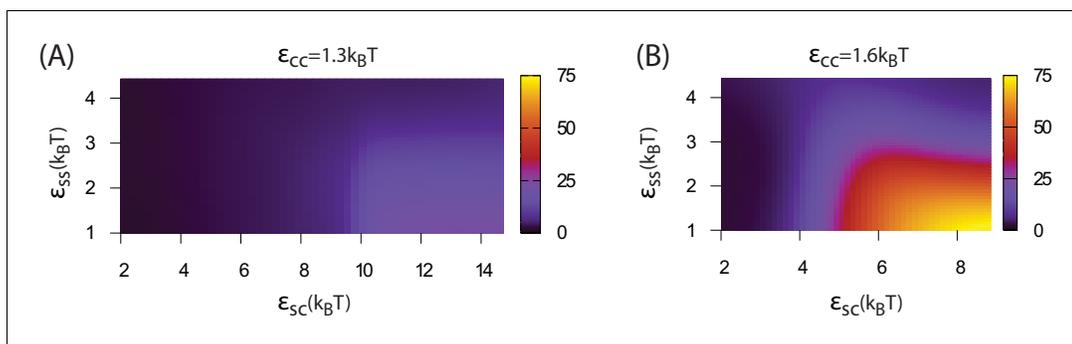
Figure 4. Dependence of assembly pathway on shell-cargo and shell-shell interaction strength. The assembly order parameter, defined as the maximum number of unassembled shell subunits adsorbed on a globule at any point during a trajectory, is shown as a function of ε_{SC} and ε_{SS} for indicated values of the cargo-cargo interaction ε_{CC}. Large numbers of adsorbed unassembled subunits (>32) indicate the two step assembly mechanism (**Figure 2A**), whereas smaller values correspond to simultaneous assembly and cargo condensation (**Figure 2B**).
DOI: 10.7554/eLife.14078.016

The following figure supplement is available for figure 4:

Figure supplement 1. Assembly order parameter values for $\varepsilon_{CC} = 1.8$ and $\varepsilon_{CC} = 2.0$.
DOI: 10.7554/eLife.14078.017

on shell-shell interaction strengths, since multiple nucleation events were rare in the carboxysome assembly experiments (*Cameron et al., 2013*) (however, we discuss potential complicating factors within the cellular environment below). To quantify the relationship between assembly mechanism and parameter values, we calculate an assembly order parameter, defined as the maximum number of unassembled subunits adsorbed onto a globule during an assembly trajectory. The order parameter is shown as a function of the interaction strengths in **Figure 4**. For $\varepsilon_{CC} \geq 1.6$ and $\varepsilon_{SS} \leq 3$ we observe large values of the order parameter (*e.g.* >32, the red and yellow regions in **Figure 4**), which indicate formation of a large amorphous globule consient with the procarboxysome precursor to carboxysome shell assembly (*Cameron et al., 2013*).

Other assembly products

Outside of the optimal parameter ranges, we observe several classes of alternative outcomes. Overly weak shell-shell interactions fail to drive assembly. For $\varepsilon_{CC} = 1.6$ and $\varepsilon_{SC} \leq 4$ the cargo vapor phase is metastable, and the system remains 'Unnucleated' (with no cargo globule) on simulated timescales (we discuss alternative initial conditions below). Stronger cargo-cargo or shell-cargo interactions result in unassembled 'Globules', where a cargo globule forms but the shell subunits on its surface fail to nucleate. As ε_{SS} increases, we observe assembly on the globule, leading either to complete shells or two classes of incomplete assembly. In the first incomplete case, 'Attached', one or more shells almost reaches completion, but fails to detach from the droplet within simulated timescales. 'Attached' configurations occur for low ε_{SC}, when the subunit-cargo interaction does not provide a strong enough driving force for the last subunit(s) to penetrate the cargo and close the shell. Overly strong interactions drive the other class of incomplete assembly: 'Over-nucleated/Malformed', in which an excess of partially assembled shells deplete the system of free subunits before any shells are completed. In this regime it is also common to observe malformed structures, in which defects become trapped within growing shells.

As the cargo-cargo interaction increases ($\varepsilon_{CC} \geq 1.8$), multiple effects narrow the parameter range that leads to complete assembly and detachment. Firstly, cargo globules nucleate rapidly at multiple locations within the simulation box, increasing the likelihood of the 'Over-nucleated' outcome. Secondly, the threshold value of ε_{SC} required for cargo penetration increases, resulting in 'Attached' shells over a wider parameter range. We also observe a configuration we refer to as 'Stalled', in which shell assembly fails to penetrate the globule surface (and thus does not even proceed to the attached stage). The latter is especially prevalent for $\varepsilon_{CC} = 3.0$, when the cargo crystallizes even in the absence of shell encapsulation. For both 'Attached' and 'Stalled' configurations, regardless of the initial number of nucleation events, we typically observe coarsening into a large globule.
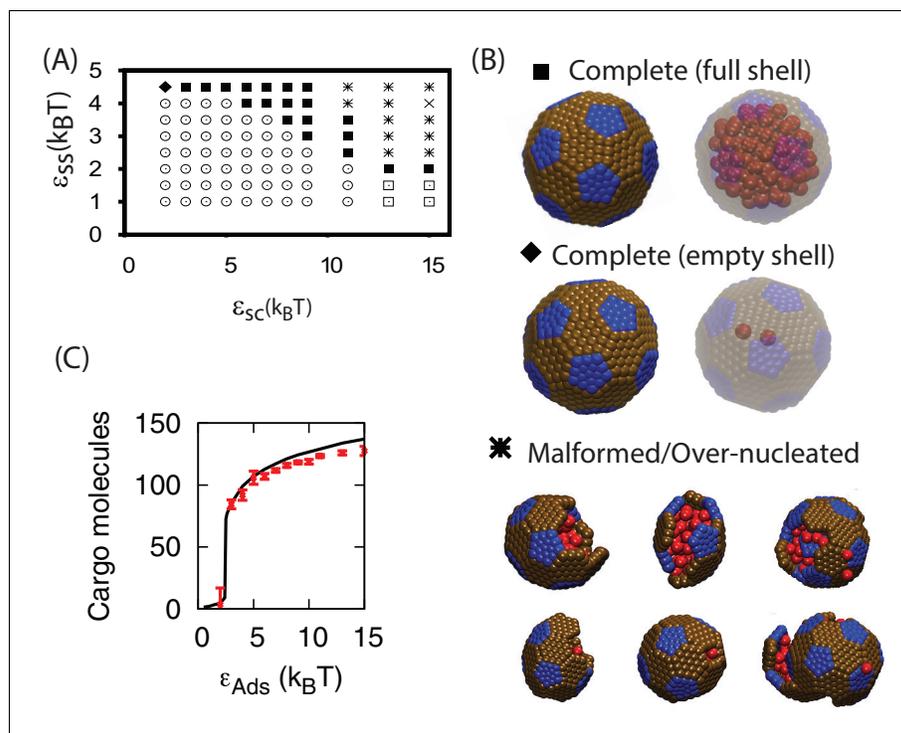
Figure 5. Results of assembly around a cargo with weak interactions ($\varepsilon_{CC} = 1.3k_BT$). (**A**) The most frequently observed assembly outcome as a function of ε_{SS} and ε_{SC}. The distribution of outcomes for $\varepsilon_{SS}=4$ is shown in *Figure 3—figure supplement 2*, and a data file containing the outcome for each trial at each parameter set is included (*Figure 3—source data 1*). (**B**) Representative snapshots for the outcomes shown in (**A**). The complete shell outcomes are shown with the excluders rendered opaque (left) and transparent (right) to enable visualizing the encapsulated cargo. (**C**) The number of cargo molecules encapsulated by shells assembled in dynamics simulations (red symbols) is compared to the results of equilibrium simulations (black line). The dynamics results are averaged over all complete shells (for any ε_{SS}) assembled at each value of ε_{SC}, the error bars indicate 95% confidence intervals. Most simulations were performed for 3×10^8 timesteps; simulations with $\varepsilon_{SS}=4.5$, $\varepsilon_{SC} \leq 4$, and $\varepsilon_{CC} = 1.3$ exhibited partially assembled shells at 3×10^8 timesteps, and were continued up to 7.2×10^9 timesteps.

DOI: 10.7554/eLife.14078.018

The following figure supplements are available for figure 5:

Figure supplement 1. Assembly yields calculated by simulation and theory.
DOI: 10.7554/eLife.14078.019

Figure supplement 2. The effect of varying cargo diameter on assembly.
DOI: 10.7554/eLife.14078.020

Simultaneous shell assembly and cargo condensation

For $\varepsilon_{CC} = 1.3$ the cargo forms an equilibrium vapor phase in the absence of shell subunits. However, above threshold values of ε_{SS} and ε_{SC}, the diffuse cargo molecules drive nucleation of shell assembly. The subsequent assembly pathway depends sensitively on the shell-cargo interaction strength. For low ε_{SC} (*Figure 2C*), assembly captures only a few cargo molecules, leading to complete, but nearly empty shells. For larger ε_{SC} (*Figure 2B*, and Simulation *Video 2*), the shell-cargo interactions drive local condensation of cargo molecules. Shell assembly and cargo complexation then proceed in concert, resembling the mechanism proposed for assembly of α-carboxysomes (*Iancu et al., 2010*). Thus, tuning the shell-cargo interaction dramatically affects cargo loading, with a sharp transition from empty to filled shells around $\varepsilon_{SC}=2$. This transition closely tracks the equilibrium filling fraction (*Figure 5C*), measured by simulating a complete shell made permeable to cargo molecules. This effect is comparable to the condensation of water vapor below its dew point inside of hydrophilic cavities. In contrast, assembly around a globule only generates full shells.

Assembly of full shells (by either pathway, *Figure 2A* or *Figure 2B*) is typically about two orders of magnitude faster than assembly of empty shells (*Figure 2C*). This disparity demonstrates the key role that the cargo plays in promoting shell association, during all stages of assembly. Cargo molecules initially promote shell nucleation by stabilizing interactions among small, sub-nucleated clusters. Then, the presence of a condensed globule provides a large cross-section for adsorption of additional subunits, significantly enhancing the flux of subunits to the partial capsid, thus increasing its growth rate. The condensed cargo particularly facilitates insertion of the last few subunits, which are significantly hindered by steric interactions, as noted previously for simulations of empty virus capsids (*Nguyen et al., 2007*).

Figure 5A shows how the products of assembly around cargo with weak interactions depends on parameters. While moderate parameter values lead to complete assembly, overly weak ε_{SC} and ε_{SS} (lower left region of *Figure 5A*) prevent shell nucleation, leading to the 'Unnucleated' outcome. In the limit of large ε_{SC} but weak ε_{SS} the shell-cargo interaction stabilizes small disordered globules (~ 50 cargo particles, lower right region of *Figure 5A*), while under strong subunit and weak cargo interactions ($\varepsilon_{SS} = 4.5$, $\varepsilon_{SC} < 5$) shells nucleate but cannot condense the cargo, leading to the complete but slow assembly just discussed. As for assembly around a globule, overly strong interactions lead to overnucleation and malformed shells. However, the predominant mode of malformation is now shell collapse. Because the cargo is below its dew point, the locally condensed globule leads to a negative pressure on the shell subunits, which can flatten the shell and thus prevent closure of a symmetric shell.

Thermodynamic model

The simple free energy model (*Equations (1–2)*) reproduces the threshold parameter values required for shell assembly with no adjustable parameters (color map in *Figure 3*). Since it is an equilibrium model and only considers the free energy difference between complete and unassembled configurations, it cannot distinguish between parameter values that lead to complete assembly or kinetic traps at the long but finite simulation times. However, the thermodynamic calculation does suggest that the simulations resulting in 'Attached' shells would eventually reach completion on a longer timescale. We do not show Δf_{assem} in *Figure 5A* because the globule is always less favorable than assembled shells for $\varepsilon_{CC} = 1.3$, but the yield of well-formed shells in our simulations roughly follows the prediction of the equilibrium theory (*Figure 5—figure supplement 1*).

Effects of varying other parameters or initial conditions

To investigate whether the results described above depend on assumptions within our model, we performed several sets of additional simulations. Firstly, we performed simulations in which the ratio between cargo diameter in shell subunit size was varied. As shown in *Figure 5—figure supplement 2*, assembly is most robust for our default cargo diameter (for which the model was parameterized), but productive assembly occurs for cargo diameters varied over a factor of four. Secondly, we performed assembly simulations with anisotropic cargo molecules with a shape motivated by the octomer structure of the RuBisCO holoenzyme (*Figure 2—figure supplement 2*).

Thirdly, we performed a set of simulations in which we pre-equilibrated the cargo globule before introducing shell subunits into the system (*Figure 3—figure supplement 2*, Simulation *Video 3*). Investigating this alternative initial condition was motivated by the fact that RuBisCO is present in the cell before induction of the
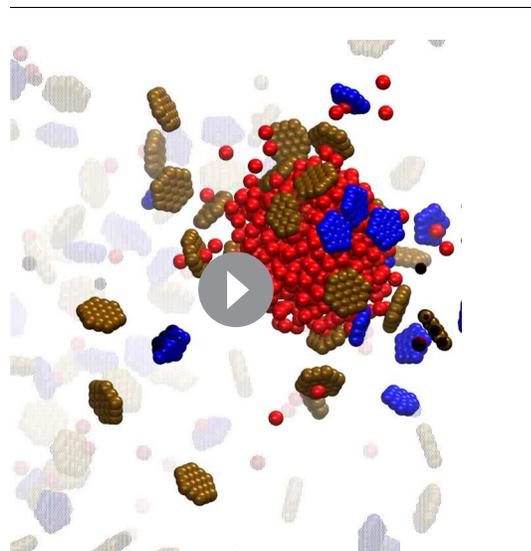
Video 3. Animation of a simulation with a pre-equilibrated cargo globule. Parameters are $\varepsilon_{CC} = 1.6$, $\varepsilon_{SC} = 6$, and $\varepsilon_{SS} = 3.5$.
DOI: 10.7554/eLife.14078.021
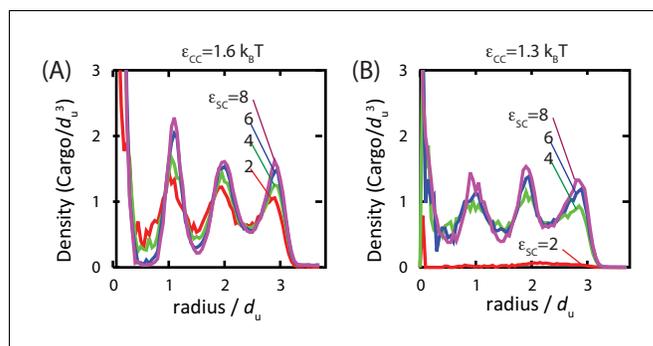
Figure 6. Order of the encapsulated cargo. The spherically averaged density of cargo molecules inside a shell is shown as a function of radius for (A) $\varepsilon_{\mathrm{CC}}=1.6$ and (B) $\varepsilon_{\mathrm{CC}}=1.3$ for indicated values of the cargo-shell adhesion strength $\varepsilon_{\mathrm{SC}}$, measured in equilibrium simulations. The density of the encapsulated cargo ranges from below random close packing to near hexagonal close packing density as $\varepsilon_{\mathrm{CC}}$ and $\varepsilon_{\mathrm{SC}}$ are increased (see *Figure 3—figure supplement 3*). A snapshot of cargo inside the shell is shown in *Figure 5—figure supplement 2*. The raw data for this figure is provided in *Figure 6—source data 1*.
DOI: 10.7554/eLife.14078.022

The following source data and figure supplements are available for figure 6:

Source data 1. Raw data for *Figure 6*.
DOI: 10.7554/eLife.14078.023

Figure supplement 1. Ordering of the encapsulated cargo is primarily driven by confinement, not adhesion to the inner surface of the shell.
DOI: 10.7554/eLife.14078.024

Figure supplement 1—source data 1. Raw data for *Figure 6—figure supplement 1*.
DOI: 10.7554/eLife.14078.025

---

carboxysome gene in the experiments of Ref. (*Cameron et al., 2013*), and by the observation that multiple carboxysomes bud sequentially in time from a single procarboxysome. For $\varepsilon_{\mathrm{CC}} = 1.6$ the results are very similar to those obtained without pre-equilibrating the cargo. However, for $\varepsilon_{\mathrm{CC}}>1.6$, successful assembly and detachment is limited to more narrow ranges of shell-shell and shell-cargo interaction strengths than in *Figure 3*, due to an increased prevalence of 'Attached' and 'Stalled' configurations. The latter are particularly common for $\varepsilon_{\mathrm{CC}}=3$, when the cargo forms a hexagonally close packed crystal which strongly resists deformation by shell protein assembly.

Taken together, the results from both assembly protocols (*Figure 3* and *Figure 3—figure supplement 2*) suggest that moderate effective cargo-cargo interactions are most consistent with the observations of shell assembly and budding in Refs. (*Cameron et al., 2013*; *Chen et al., 2013*). Such interactions are strong enough to drive cargo globule formation, but malleable enough to allow shell assembly to deform and eventually sever intra-globule interactions.

Organization of encapsulated cargo

Studies of assembled carboxysomes report varying degrees of order for the encapsulated cargo, ranging from none to paracrystalline order (*Iancu et al., 2007*; *2010*; *Kaneko et al., 2006*; *Schmid et al., 2006*). We therefore studied the relationship between cargo order and interaction parameters using equilibrium simulations (see *Figure 6* and *Figure 6—figure supplement 1*). Below $\varepsilon_{\mathrm{CC}}<3k_{\mathrm{B}}T$, we do not observe true fcc order of the encapsulated cargo. However, for all parameters leading to significant filling, even those well below the cargo liquid-vapor transition, the cargo becomes organized in concentric layers (*Figure 6*). We observe similar cargo organizations within shells which have budded from cargo globules in dynamical simulations. These results demonstrate that ordering of the cargo does not require crystallinity of the initial globule. Moreover, the magnitude of ordering increases with cargo loading, but, for fixed loading, is essentially independent of the cargo-shell interaction strength $\varepsilon_{\mathrm{SC}}$. We observe ordering within filled shells due to confinement, even if even if $\varepsilon_{\mathrm{SC}}$ is set to 0 (*Figure 6—figure supplement 1*), as previously noted by Iancu *et al.* (*Iancu et al., 2007*).

Table 1. Description of the assembly outcomes presented in *Figures 3*,*5*.

Symbol	Name	Description
■	Complete shell (full)	Complete shell, full of cargo molecules
◆	Complete shell (empty)r	Complete shell, almost empty of cargo molecules
●	Attached	Nearly complete shells attached to a globule by a neck of cargo
✳	Over-nucleated/Malformed	Multiple globules, with incomplete or malformed shells on their surfaces
×	Stalled	Large globule with multiple incomplete or malformed shells on its surface
□	Globule	Cargo globule with unassembled shell subunits on its surface
⊙	Unnucleated	Diffuse subunits and cargo molecules

Discussion

We have described an equilibrium theory and a dynamical computational model for the assembly of shells around a fluid cargo. Our simulations show that assembly can proceed by two classes of pathways: (i) a multi-step process in which the cargo forms a dense globule, followed by adsorption, assembly, and budding of shell proteins, or (ii) single-step assembly, with simultaneous aggregation of cargo molecules and shell assembly. This result demonstrates that the minimal interactions included in our model are sufficient to drive both classes of assembly pathways, suggesting that they are a generic feature of assembly around a fluid cargo. Moreover, while we cannot rule out the existence of active mechanisms in biological examples such as carboxysomes, our model demonstrates that the same interactions which drive assembly of shells can also drive budding from and closure around an amorphous globule of cargo.

Our results suggest bounds on the relative strengths of interactions that drive BMC assembly in cells. The decisive control parameter determining the assembly pathway is the cohesive energy between cargo molecules, which could arise through direct cargo-cargo interactions or be mediated by auxiliary proteins (*Cameron et al., 2013*). Relatively weak cargo interactions lead to single-step assembly pathways, while stronger interactions favor formation of the cargo-shell globule. However, the strength of cargo-shell and shell-shell interactions also play a role. Strong shell-shell interactions cause assembly to proceed rapidly during globule formation, limiting the size of the globule. Moreover, if a large globule is already present (*e.g.* due to time-dependent protein concentrations within a cell), strong interactions tend to drive malformed assemblies. We find that an important functional difference between the two classes of assembly pathways is control over the amount of packaged cargo. While the multi-step assembly pathways always generate a shell filled with cargo molecules, shells assembling around a diffuse cargo can be tuned from nearly empty to completely full by controlling the strength of cargo-shell interactions.

These results have implications for reengineering BMCs to encapsulate new core enzymes. Recent works demonstrated that protein cargos can be targeted to BMCs via encapsulation peptides that mediate cargo-shell interactions. However, packaged amounts were much lower than for native core enzymes (*Parsons et al., 2010*; *Choudhary et al., 2012*; *Lassila et al., 2014*). Our simulations show that both cargo-shell and cargo-cargo interactions (direct or mediated) must be controlled to assemble full shells.

We also find that a general equilibrium theory describes the ranges of parameter values for which assembly occurs. However, the dynamical simulations demonstrate that, at finite timescales, there is a rich variety of assembly morphologies. Formation of ordered, full shells requires a delicate balance of cargo-cargo, cargo-shell, and shell-shell interactions, all of which must be on the order $5 - 10k_{\rm B}T$. This constraint is consistent with previous studies on viruses and other assembly systems, which found that formation of ordered states requires multiple, cooperative weak interactions between subunits (*Hagan, 2014*; *Whitelam and Jack, 2015*). Outside of optimal parameter regimes, the simulations predict alternative outcomes, ranging from no assembly to various alternative trapped intermediates, with the morphology depending on which interaction is strongest. We find that assembly is least robust to parameter variations when the cargo crystallizes before shell assembly. The assembling shell is unable to deform or penetrate the cargo complex, leading to defect-riddled, non-budded complexes. Within the limits of our simplified model, this observation suggests that

procarboxysome complexes are at least partially fluid prior to successful shell assembly. Moreover, we find that observations of ordered cargo within assembled shells may be explained by packing constraints.

An important limitation of the present study is that the model interactions are specific to the shell geometry shown in *Figure 1* (containing 20 hexamers) because alternating edges on hexagonal sub-units have attractive interactions only with pentagonal subunits. In reality BMCs contain many more hexamers (formed from multiple protein sequences) and thus must include a greater range of hex-amer-hexamer interactions. Extension of the model to allow for this possibility would allow consider-ation of two important questions: (1) The mechanism controlling insertion of the 12 pentagons required for a closed shell topology. (2) The relationship between assembly pathway and BMC size polydispersity. In particular, experiments suggest that β-carboxysomes are more polydisperse than α-carboxysomes (*Price and Badger, 1991*; *Shively et al., 1973*; *Shively et al., 1973*; *Iancu et al., 2007*; *2010*; *Kerfeld et al., 2010*; *Tanaka et al., 2008*). We speculate that in the case of assembly around vapor-phase cargo, the size of the assembling shell will be primarily dictated by the pre-ferred shell protein curvature and thus relatively uniform. However, during assembly around a con-densed globule, the shell protein interactions could be strained to accommodate a globule which is larger or smaller than the preferred curvature, causing the shell size to depend on a complex bal-ance of intermolecular interaction strengths and variables such as the local RuBisCO concentration.

Our model is minimal, intended to elucidate general principles of assembly around a fluid cargo, and thus may apply to diverse systems including prokaryotic microcompartments, viruses, and engi-neered delivery vehicles. The predicted trends for how assembly mechanisms and morphologies vary with control parameters can be experimentally tested by microscopy experiments. Such testing will be most straightforward in vitro (*e.g. Luque et al., 2014*; *Douglas and Young, 1998*; *Rurup et al., 2014*; *Patterson et al., 2014*; *Patterson et al., 2012*; *Zhu et al., 2014*; *Rhee et al., 2011*; *Rurup et al., 2014*; *Wörsdörfer et al., 2012*), where subunit-subunit interactions can be tuned by varying solution conditions and the stoichiometries of shell and cargo species can be read-ily varied. While there is currently no BMC assembly system starting from purified components, our findings can be tested in vivo by mutations which alter known protein binding interfaces, or by alter-ing expression levels of RuBisCO or carboxysome proteins.

We anticipate that our model can serve as a qualitative guide for understanding how such multi-component complexes assemble in natural systems, or to reengineer them for new applications. More broadly, our results demonstrate that the properties of encapsulated cargo, such as its topol-ogy, geometry and interaction strengths, strongly influence assembly pathways and morphologies.

Materials and methods

Computational model

Shell subunits

We have adapted a model for virus assembly (*Perlmutter et al., 2013*; *2014*; *Perlmutter and Hagan, 2015a*; *Wales, 2005*; *Fejer et al., 2009*; *Johnston et al., 2010*; *Ruiz-Herrero and Hagan, 2015*) to describe assembly of an icosahedral shell around a fluid cargo. Each subunit contains 'Attractors' on its perimeter that mediate subunit-subunit attractions (as in *Ruiz-Herrero and Hagan, 2015*). Attractor interactions are specific – complementary pairs of Attractors (see *Figure 1A,B* and appendix 1) have short-range interactions (modeled by a Morse potential), whereas non-comple-mentary pairs have no interactions. A repulsive interaction between pairs of 'Top' (type 'T') pseudoa-toms favors the correct subunit-subunit angle. The 'Bottom' (type 'B') pseudoatoms mediate short-ranged subunit-cargo attractions (*e.g.* due to interactions with shell 'encapsulation peptides' (*Kinney et al., 2012*; *Cameron et al., 2013*; *Fan et al., 2010*)), represented by a Morse potential. We also add a layer of 'Excluders' in the plane of the 'Top' pseudoatoms, which represent subunit-cargo excluded volume interactions. The strengths of subunit-subunit and subunit-cargo attractions are parameterized by potential well depths ε_{SS} and ε_{SC} respectively (appendix 1).

Cargo

As a minimal representation of globular proteins, the cargo is modeled as spherical particles which interact via an attractive Lennard-Jones (LJ) potential, with well-depth ε_{CC}. The attractions implicitly

model hydrophobic and screened electrostatics interactions between cargo molecules, as well as effective cargo-cargo interactions mediated by auxiliary proteins (*e.g.* the carboxysome protein CcmM (*Cameron et al., 2013*)).

Simulations

We simulated assembly dynamics using the Langevin dynamics algorithm in HOOMD (a software package that uses GPUs to perform highly efficient dynamics simulations [*Anderson et al., 2008*]) and periodic boundary conditions to represent a bulk system. The subunits are modeled as rigid bodies (*Nguyen et al., 2011*). The simulations were performed using a set of fundamental units (URL. http://codeblue.umich.edu/hoomd-blue/doc/page_units.html), with $1d_u$ defined as the circum-radius of the pentagonal subunit (the cargo diameter is also set to $1\ d_u$). Unless specified otherwise, each simulation contained enough subunits to form four complete shells (48 pentamers and 80 hexamers) and 611 cargo particles (a shell typically encapsulates 120–130 cargo particles) in a cubic box with side length $40d_u$. The simulation time step was 0.001 in dimensionless time units, and dynamics was performed for 3×10^8 timesteps unless mentioned otherwise.

We performed two sets of simulations, using different initial conditions. In the first, simulations were initialized by introducing cargo particles and shell subunits simultaneously with random positions and orientations (except avoiding high-energy overlaps). The second set of initial conditions was motivated by the possibility that the cargo globule could form before shell subunits reach sufficient concentrations within the cell to undergo assembly. To model this situation, we pre-equilibrated the cargo by performing a long simulation with only cargo particles present. Shell subunits were then introduced with random positions and orientations (excluding high-energy overlaps). For $\varepsilon_{CC} \geq 1.6$, the assembly simulations thus began with a cargo globule already present. For $\varepsilon_{CC} < 1.6$ the two protocols are equivalent, since no globule forms during cargo equilibration.

Sample sizes

To cover the largest range of parameter space possible given the computational expense associated with each simulation, we performed 5 independent simulations at most parameter sets. To assess statistical error and to estimate the distribution of different assembly outcomes, we performed 10 independent trials for one value of ε_{SS} at each value of ε_{SC} and ε_{CC}. We also performed additional simulations at parameter sets for which 5 trials did not result in a majority outcome, or when necessary to obtain better statistics on the number of encapsidated cargo particles. Based on these results, performing additional simulations at other parameter values would not qualitatively change our results. (It would increase the statistical accuracy of estimated boundaries between different outcomes; however, these boundaries correspond to crossovers rather than sharp transitions.)

Thermodynamics of assembly around a fluid cargo

To complement the finite-time simulations, we have developed a general thermodynamic description of assembly around a fluid cargo. We consider shells composed of species $\alpha = 1, 2, \ldots M$, with n_α^{shell} subunits of species α in a complete shell, which encapsulates n_0 cargo molecules (the index 0 refers to cargo molecules henceforth). Assembly occurs from a dilute solution of cargo molecules with density ρ_0, shell subunits with density ρ_α for each species, and the density of assembled, full shells as ρ_{shell}. These are in equilibrium with a globule containing n_0^{glob} cargo molecules and n_α^{glob} subunits for each species α. We assume that, due to the asymmetric nature of the shell-cargo interaction, the shell subunits reside at the exterior of the globule (as we observe in our simulations). The globule containing unassembled shell subunits thus resembles a spherical microemulsion droplet (*Safran, 1994*). Minimizing the total free energy (see appendix 2) gives:

$$v_0 \rho_{shell} = \exp\left[-\left(G_{shell} - \sum_\alpha n_\alpha^{shell} \mu_\alpha\right)/k_B T\right] \quad (1)$$

where G_{shell} is the interaction free energy of the assembled shell and μ_α are the chemical potentials of free cargo molecules and shell subunits, given by $\mu_\alpha = k_B T \ln(\rho_\alpha v_0)$, with v_0 a standard state volume and the globule composition given by

$$\frac{\partial G_{\text{glob}}(\{n_\alpha^{\text{glob}}\})}{\partial n_\alpha^{\text{glob}}} = \mu_\alpha \quad \text{for } \alpha = 0 \dots M, \tag{2}$$

with $G_{\text{glob}}(n_{\text{s}}^{\text{glob}}, n_0^{\text{glob}})$ as the globule free energy.

(1) – (2) are the general equilibrium description for a system of assembling shells with a disordered-phase intermediate; application to a specific system requires specifying the forms of G_{shell} and G_{glob}. In appendix 2 we specify these equations for our computational model, allowing us to compare the equilibrium calculation with simulation results, using no free parameters.

To compare the relative stabilities of the globule and assembled shells, we also calculate the free energy difference

$$\Delta f_{\text{assem}} = f_{\text{tot}}(\{n_\alpha^{\text{glob}} = 0\}) - f_{\text{tot}}(\rho_{\text{shell}} = 0), \tag{3}$$

where the first term on the right-hand side is the minimized free energy for a system containing shells and free subunits but no globule, while the second term corresponds to the minimized free energy for a system containing subunits and the globule, but no assembly.

Acknowledgements

Additional information

Funding

Author contributions

JDP, MFH, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; FM, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article

Author ORCIDs

Michael F Hagan, http://orcid.org/0000-0002-9211-2434

References

Anderson JA, Lorenz CD, Travesset A. 2008. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**:5342–5359. doi: 10.1016/j.jcp.2008.01.047

Axen SD, Erbilgin O, Kerfeld CA. 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Computational Biology* **10**:e1003898. doi: 10.1371/journal.pcbi.1003898

Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC. 1999. The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation. *Journal of Bacteriology* **181**:5967–5975.

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA, Savage DF. 2012. Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences of the United States of America* **109**:478–483. doi: 10.1073/pnas.1108557109

Borodavka A, Tuma R, Stockley PG. 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences of the United States of America* **109**:15769–15774. doi: 10.1073/pnas.1204357109

Cadena-Nava RD, Comas-Garcia M, Garmann RF, Rao AL, Knobler CM, Gelbart WM. 2012. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *Journal of Virology* **86**:3318–3326. doi: 10.1128/JVI.06566-11

Cai F, Dou Z, Bernstein SL, Leverenz R, Williams EB, Heinhorst S, Shively J, Cannon GC, Kerfeld CA. 2015. Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life* **5**:1141. doi: 10.3390/life5021141

Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. 2013. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**:1131–1140. doi: 10.1016/j.cell.2013.10.044

Chen AH, Robinson-Mosher A, Savage DF, Silver PA, Polka JK. 2013. The Bacterial Carbon-Fixing Organelle Is Formed by Shell Envelopment of Preassembled Cargo. *PLoS ONE* **8**:e76127. doi: 10.1371/journal.pone.0076127

Choudhary S, Quin MB, Sanders MA, Johnson ET, Schmidt-Dannert C. 2012. Engineered protein nano-compartments for targeted enzyme localization. *PloS One* **7**:e33342. doi: 10.1371/journal.pone.0033342

Comas-Garcia M, Cadena-Nava RD, Rao AL, Knobler CM, Gelbart WM. 2012. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *Journal of Virology* **86**:12271–12282. doi: 10.1128/JVI.01695-12

Comas-Garcia M, Garmann RF, Singaram SW, Ben-Shaul A, Knobler CM, Gelbart WM. 2014. Characterization of Viral Capsid Protein Self-Assembly around Short Single-Stranded RNA. *The Journal of Physical Chemistry. B.* doi: 10.1021/jp503050z

Devkota B, Petrov AS, Lemieux S, Boz MB, Tang L, Schneemann A, Johnson JE, Harvey SC. 2009. Structural and electrostatic characterization of pariacoto virus: implications for viral assembly. *Biopolymers* **91**:530–538. doi: 10.1002/bip.21168

Dixit SK, Goicochea NL, Daniel MC, Murali A, Bronstein L, De M, Stein B, Rotello VM, Kao CC, Dragnea B. 2006. Quantum dot encapsulation in viral capsids. *Nano Letters* **6**:1993–1999. doi: 10.1021/nl061165u

Douglas T, Young M. 1998. Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **393**:152–155. doi: 10.1038/30211

Dykeman EC, Stockley PG, Twarock R. 2013. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of Molecular Biology* **425**:3235–3249. doi: 10.1016/j.jmb.2013.06.005

Dykeman EC, Stockley PG, Twarock R. 2014. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proceedings of the National Academy of Sciences of the United States of America* **111**:5361–5366. doi: 10.1073/pnas.1319479111

Elrad OM, Hagan MF. 2010. Encapsulation of a polymer by an icosahedral virus. *Physical Biology* **7**:045003. doi: 10.1088/1478-3975/7/4/045003

Erbilgin O, McDonald KL, Kerfeld CA. 2014. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Applied and Environmental Microbiology* **80**. doi: 10.1128/AEM.03887-13

Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA. 2010. Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences of the United States of America* **107**:7509–7514. doi: 10.1073/pnas.0913199107

Faustino AF, Carvalho FA, Martins IC, Castanho MARB, Mohana-Borges R, Almeida FCL, Da Poian AT, Santos NC. 2014. Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomedicine: Nanotechnology, Biology and Medicine* **10**:247–255. doi: 10.1016/j.nano.2013.06.004

Fejer SN, James TR, Hernández-Rojas J, Wales DJ. 2009. Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. *Physical Chemistry Chemical Physics* **11**:2098–2104. doi: 10.1039/b818062h

Garmann RF, Comas-Garcia M, Gopal A, Knobler CM, Gelbart WM. 2014a. The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *Journal of Molecular Biology* **426**:1050–1060. doi: 10.1016/j.jmb.2013.10.017

Garmann RF, Comas-Garcia M, Koay MS, Cornelissen JJ, Knobler CM, Gelbart WM. 2014b. Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *Journal of Virology* **88**. doi: 10.1128/JVI.01044-14

Gelbart WM, Ben-Shaul A, Roux D. 1994. *Micelles, Membranes, Microemulsions, and Monolayers.* New York, NY: Springer New York. doi: 10.1007/978-1-4613-8389-5

Hagan MF, Chandler D. 2006. Dynamic pathways for viral capsid assembly. *Biophysical Journal* **91**:42–54. doi: 10.1529/biophysj.105.076851

Hagan MF. 2008. Controlling viral capsid assembly with templating. *Physical Review E* **77**. doi: 10.1103/PhysRevE.77.051904

Hagan MF. 2009. A theory for viral capsid assembly around electrostatic cores. *The Journal of Chemical Physics* **130**:114902. doi: 10.1063/1.3086041

Hagan MF, Elrad OM, Jack RL. 2011. Mechanisms of kinetic trapping in self-assembly and phase transformation. *The Journal of Chemical Physics* **135**:104115. doi: 10.1063/1.3635775

Hagan MF. 2014. Modeling Viral Capsid Assembly. *Advances in Chemical Physics* **155**:1–68. doi: 10.1002/9781118755815.ch01

Hu T, Shklovskii BI. 2007. Kinetics of viral self-assembly: Role of the single-stranded RNA antenna. *Physical Review E* **75**. doi: 10.1103/PhysRevE.75.051901

Iancu CV, Ding HJ, Morris DM, Dias DP, Gonzales AD, Martino A, Jensen GJ. 2007. The structure of isolated Synechococcus strain WH8102 carboxysomes as revealed by electron cryotomography. *Journal of Molecular Biology* **372**:764–773. doi: 10.1016/j.jmb.2007.06.059

Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. 2010. Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *Journal of Molecular Biology* **396**:105–117. doi: 10.1016/j.jmb.2009.11.019

Johnson JM, Willits DA, Young MJ, Zlotnick A. 2004. Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. *Journal of Molecular Biology* **335**:455–464. doi: 10.1016/j.jmb.2003.10.059

Johnston IG, Louis AA, Doye JPK. 2010. Modelling the self-assembly of virus capsids. *Journal of Physics: Condensed Matter* **22**:104101. doi: 10.1088/0953-8984/22/10/104101

Kaneko Y, Danev R, Nagayama K, Nakamoto H. 2006. Intact carboxysomes in a cyanobacterial cell visualized by hilbert differential contrast transmission electron microscopy. *Journal of Bacteriology* **188**:805–808. doi: 10.1128/JB.188.2.805-808.2006

Kerfeld CA, Heinhorst S, Cannon GC. 2010. Bacterial microcompartments. *Annual Review of Microbiology* **64**:391–408. doi: 10.1146/annurev.micro.112408.134211

Kerfeld CA, Erbilgin O. 2015. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in Microbiology* **23**:22–34. doi: 10.1016/j.tim.2014.10.003

Kickhoefer VA, Rajavel KS, Scheffer GL, Dalton WS, Scheper RJ, Rome LH. 1998. Vaults are up-regulated in multidrug-resistant cancer cell lines. *The Journal of Biological Chemistry* **273**:8971–8974. doi: 10.1074/jbc.273.15.8971

Kinney JN, Salmeen A, Cai F, Kerfeld CA. 2012. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *The Journal of Biological Chemistry* **287**:17729–17736. doi: 10.1074/jbc.M112.355305

Kivenson A, Hagan MF. 2010. Mechanisms of capsid assembly around a polymer. *Biophysical Journal* **99**:619–628. doi: 10.1016/j.bpj.2010.04.035

Lassila JK, Bernstein SL, Kinney JN, Axen SD, Kerfeld CA. 2014. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *Journal of Molecular Biology* **426**:2217—28.. doi: 10.1016/j.jmb.2014.02.025

Lindenbach BD, Rice CM. 2013. The ins and outs of hepatitis C virus entry and assembly. *Nature Reviews. Microbiology* **11**:688–700. doi: 10.1038/nrmicro3098

Luque D, Escosura Andrés de la, Snijder J, Brasch M, Burnley RJ, Koay MST, Carrascosa JL, Wuite GJL, Roos WH, Heck AJR, Cornelissen JJLM, Torres T, Castón JR. 2014. Self-assembly and characterization of small and monodisperse dye nanospheres in a protein cage. *Chem. Sci.* **5**:575–581. doi: 10.1039/C3SC52276H

Mahalik JP, Muthukumar M. 2012. Langevin dynamics simulation of polymer-assisted virus-like assembly. *The Journal of Chemical Physics* **136**:135101. doi: 10.1063/1.3698408

Malyutin AG, Dragnea B. 2013. Budding pathway in the templated assembly of viruslike particles. *The Journal of Physical Chemistry. B* **117**. doi: 10.1021/jp405603m

Mecke M, Winkelmann J, Fischer J. 1997. Molecular dynamics simulation of the liquid–vapor interface: The lennard-jones fluid. *J. Chem. Phys* **107**:9264–9270 . doi: 10.1063/1.475217

Nguyen HD, Reddy VS, Brooks CL. 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Letters* **7**:338–344. doi: 10.1021/nl062449h

Nguyen TD, Phillips CL, Anderson JA, Glotzer SC. 2011. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. *Computer Physics Communications* **182**:2307–2313. doi: 10.1016/j.cpc.2011.06.005

Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, Warren MJ. 2010. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular Cell* **38**:305–315. doi: 10.1016/j.molcel.2010.04.008

Patel N, Dykeman EC, Coutts RH, Lomonossoff GP, Rowlands DJ, Phillips SE, Ranson N, Twarock R, Tuma R, Stockley PG. 2015. Revealing the density of encoded functions in a viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* **112**:2227–2232. doi: 10.1073/pnas.1420812112

Patterson DP, Prevelige PE, Douglas T. 2012. Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22. *ACS Nano* **6**:5000–5009. doi: 10.1021/nn300545z

Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T. 2014. Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle. *ACS Chemical Biology* **9**:359–365. doi: 10.1021/cb4006529

Perlmutter JD, Qiao C, Hagan MF. 2013. Viral genome structures are optimal for capsid assembly. *eLife* **2**:e00632.. doi: 10.7554/eLife.00632

Perlmutter JD, Perkett MR, Hagan MF. 2014. Pathways for virus assembly around nucleic acids. *Journal of Molecular Biology* **426**. doi: 10.1016/j.jmb.2014.07.004

Perlmutter JD, Hagan MF. 2015a. The Role of Packaging Sites in Efficient and Specific Virus Assembly. *Journal of Molecular Biology* **427**:2451–2467. doi: 10.1016/j.jmb.2015.05.008

Perlmutter JD, Hagan MF. 2015b. Mechanisms of virus assembly. *Annual Review of Physical Chemistry* **66**:217–239. doi: 10.1146/annurev-physchem-040214-121637

Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, Romashko I, Deshpande S, Haas K, Alvelo-Maurosa JG, Wardman C, Schnell DJ, Leschine SB, Blanchard JL. 2013. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PloS One* **8**:e54337. doi: 10.1371/journal.pone.0054337

Pfeifer F. 2012. Distribution, formation and regulation of gas vesicles. *Nature Reviews. Microbiology* **10**:705–715. doi: 10.1038/nrmicro2834

Price GD, Badger MR. 1991. Evidence for the role of carboxysomes in the cyanobacterial CO 2 -concentrating mechanism. *Canadian Journal of Botany* **69**:963–973 . doi: 10.1139/b91-124

Rhee JK, Hovlid M, Fiedler JD, Brown SD, Manzenrieder F, Kitagishi H, Nycholat C, Paulson JC, Finn MG. 2011. Colorful virus-like particles: fluorescent protein packaging by the Qβ capsid. *Biomacromolecules* **12**:3977–3981. doi: 10.1021/bm200983k

Ruiz-Herrero T, Hagan MF. 2015. Simulations show that virus assembly and budding are facilitated by membrane microdomains. *Biophysical Journal* **108**:585–595. doi: 10.1016/j.bpj.2014.12.017

Rurup WF, Snijder J, Koay MS, Heck AJ, Cornelissen JJ. 2014. Self-sorting of foreign proteins in a bacterial nanocompartment. *Journal of the American Chemical Society* **136**:3828–3832. doi: 10.1021/ja410891c

Rurup WF, Verbij F, Koay MS, Blum C, Subramaniam V, Cornelissen JJ. 2014. Predicting the loading of virus-like particles with fluorescent proteins. *Biomacromolecules* **15**:558–563. doi: 10.1021/bm4015792

Safran S. 1994. *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes*. Addison-Wesley Pub.

Schmid MF, Paredes AM, Khant HA, Soyer F, Aldrich HC, Chiu W, Shively JM. 2006. Structure of Halothiobacillus neapolitanus carboxysomes by cryo-electron tomography. *Journal of Molecular Biology* **364**:526–535. doi: 10.1016/j.jmb.2006.09.024

Shively JM, Ball F, Brown DH, Saunders RE. 1973. Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacillus neapolitanus. *Science* **182**:584–586. doi: 10.1126/science.182.4112.584

Shively JM, Ball FL, Kline BW. 1973. Electron microscopy of the carboxysomes (polyhedral bodies) of Thiobacillus neapolitanus. *Journal of Bacteriology* **116**:1405–1411.

Shively JM, Bradburne CE, Aldrich HC, Bobik TA, Mehlman JL, Jin S, Baker SH. 1998. Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes - polyhedral bodies. *Canadian Journal of Botany* **76**:906–916 . doi: 10.1139/b98-088

Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology* **15**:939–947. doi: 10.1038/nsmb.1473

Sutter M, Faulkner M, Aussignargues C, Paasch BC, Barrett S, Kerfeld CA, Liu LN. 2016. Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Letters* **16**:1590–1595. doi: 10.1021/acs.nanolett.5b04259

Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO. 2008. Atomic-Level Models of the Bacterial Carboxysome Shell. *Science* **319**:1083–1086 . doi: 10.1126/science.1151458

Wales DJ. 2005. The energy landscape as a unifying theme in molecular science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **363**:357–377. doi: 10.1098/rsta.2004.1497

Whitelam S, Jack RL. 2015. The statistical mechanics of dynamic pathways to self-assembly. *Annual Review of Physical Chemistry* **66**:143–163. doi: 10.1146/annurev-physchem-040214-121215

Wörsdörfer B, Pianowski Z, Hilvert D. 2012. Efficient in vitro encapsulation of protein cargo by an engineered protein container. *Journal of the American Chemical Society* **134**:909–911. doi: 10.1021/ja211011k

Zhang Y, Inks ES, Zhu M, Chou CJ, Fang H, Li M, Shen Y, Yi F, Xu W. 2013. Discovery of a Pair of Diastereomers as Potent HDACs Inhibitors: Determination of Absolute Configuration, Biological Activity Comparison and Computational Study. *RSC Advances* **3**:25258–25267. doi: 10.1039/c3ra43249a

Zhang R, Linse P. 2013. Icosahedral capsid formation by capsomers and short polyions. *The Journal of Chemical Physics* **138**:154901. doi: 10.1063/1.4799243

Zhu Y, Wang F, Zhang C, Du J. 2014. Preparation and mechanism insight of nuclear envelope-like polymer vesicles for facile loading of biomacromolecules and enhanced biocatalytic activity. *ACS Nano* **8**:6644–6654. doi: 10.1021/nn502386j

Zlotnick A, Porterfield JZ, Wang JC. 2013. To build a virus on a nucleic acid substrate. *Biophysical Journal* **104**:1595–1604. doi: 10.1016/j.bpj.2013.02.005

Appendix 1: Model Details

1.1 Interaction potentials

Our subunit model is based on a model for viral capsid assembly, developed by Wales (**Wales, 2005**) and Johnston et al. (**Johnston et al., 2010**), which we have adapted to describe interactions with cargo molecules.

Each subunit contains 'Attractors' on its perimeter that mediate subunit-subunit attraction (as in [**Ruiz-Herrero and Hagan, 2015**]). Attractor interactions are specific – complementary pairs of Attractors have short-range interactions (modeled by a Morse potential), whereas non-complementary pairs have no interactions. For simplicity, complementarity is defined based only on the low-energy structure (**Figure 1D**); i.e., there is no attraction between pairs of pentagons. Complementary pairs of attractors are: for the hexagon-hexagon interaction, A4-A4, A5-A6, and for the hexagon-pentagon interaction A1-A4, A2-A8, A3-A7. The strength of attractive interactions is parameterized by the well-depth ε_{SS}. Because vertex attractors (A1, A4) have multiple partners in an assembled structure, whereas edge attractors have only one, the well-depth for A1-A4 and A4-A4 interactions is set to $\varepsilon_{\text{SS}}/2$, while all other attractor interactions use ε_{SS}. The 'Top' height, or distance out of the attractor plane, sets the Top-Top distance between interacting subunits, which determines the preferred subunit-subunit angle. We use a height of $h = 1/2r_{\text{b}}$, with $r_{\text{b}} = 1$ the distance between a vertex attractor and the center of the pentagon. The 'Bottom' (type 'B') pseudoatoms mediate subunit-cargo attractions, represented by a Morse potential with well-depth ε_{SC}. We also add a layer of 'Excluders' in the plane of the 'Top' pseudoatoms (positioned as in **Figure 1**), which represent subunit-cargo excluded volume interactions.

In our model, all potentials can be decomposed into pairwise interactions. Potentials involving container subunits further decompose into pairwise interactions between their constituent building blocks – the excluders, attractors, 'Top', and 'Bottom' pseudoatoms. It is convenient to state the total energy of the system as the sum of three terms, involving subunit-subunit (U_{SS}), cargo-cargo (U_{LJ}), and subunit-cargo (U_{Ads}) interactions, each summed over all pairs of the appropriate type:

$$U = \sum_{\text{sub } i} \sum_{\text{sub } j<i} U_{\text{SS}} + \sum_{\text{cargo } i} \sum_{\text{cargo } j<i} U_{\text{LJ}} + \sum_{\text{sub } i} \sum_{\text{cargo } j} U_{\text{Ads}} \tag{A1}$$

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

Subunit-subunit interactions

The subunit-subunit 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. Thus, 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 $\{\mathbf{t}_i\}$, and 'Bottom' position $\{\mathbf{b}_i\}$. The subunit-subunit interaction potential between two subunits i and j is then defined as:

$$
\begin{aligned}
U_{\mathrm{SS}}(\{\mathbf{a}_{i\alpha}\}, \mathbf{t}_i, \mathbf{a}_j, \mathbf{t}_j) &= \varepsilon_{\mathrm{SS}}\mathbf{Ł}\big(|\mathbf{t}_i - \mathbf{t}_j|,\, \sigma_{\mathrm{t},ij}\big) \\
&+ \varepsilon_{\mathrm{SS}}\mathbf{Ł}\big(|\mathbf{b}_i - \mathbf{b}_j|,\, \sigma_{\mathrm{b}}\big) \\
&+ \varepsilon_{\mathrm{SS}}\mathbf{Ł}\big(|\mathbf{b}_i - \mathbf{t}_j|,\, \sigma_{\mathrm{tb}}\big) \\
&+ \sum_{\alpha,\beta}^{N_{ai},N_{aj}} \varepsilon_{\mathrm{SS}}\mathcal{M}\big(|\mathbf{a}_{i\alpha} - \mathbf{a}_{j\beta}|,\, r_0,\varrho, r_{\mathrm{cut}}^{\mathrm{att}}\big)
\end{aligned}
\tag{A2}
$$

where $\varepsilon_{\mathrm{SS}}$ is an adjustable parameter which both sets the strength of the subunit-subunit attraction at each attractor site and scales the repulsive interactions which enforce the geometry, N_{ai} is the number of attractor pseudoatoms in subunit i, $\sigma_{\mathrm{tb}} = 1.8 r_{\mathrm{b}}$ is the diameter of the 'Top' - 'Bottom' interaction (this prevents subunits from binding in inverted configurations (**Johnston et al., 2010**), and $\sigma_{\mathrm{b}} = 1.5 r_{\mathrm{b}}$ is the diameter of the 'Bottom' - 'Bottom' interaction.

In contrast to the latter parameters, $\sigma_{\mathrm{t},ij}$ the effective diameter of the 'Top' - 'Top' interaction, depends on the species of subunits i and j; denoting a pentagonal or hexagonal subunit as p or h respectively, $\sigma_{\mathrm{t,pp}} = 2.1 r_{\mathrm{b}}$, $\sigma_{\mathrm{t,hh}} = 2.436 r_{\mathrm{b}}$, and $\sigma_{\mathrm{t,ph}} = 2.269 r_{\mathrm{b}}$. The parameter r_0 is the minimum energy attractor distance, set to $0.2 r_{\mathrm{b}}$, ϱ is a parameter determining the width of the attractive interaction, set to $4 r_{\mathrm{b}}$, and $r_{\mathrm{cut}}^{\mathrm{att}}$ is the cutoff distance for the attractor potential set to $2.0 r_{\mathrm{b}}$. Since the interactions just described are sufficient to describe assembly of the shell subunits, we included no excluder-excluder interactions.

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

$$
\mathbf{Ł}(x,\sigma) \equiv \theta(\sigma - x)\left[\left(\frac{\sigma}{x}\right)^{12} - 1\right]
\tag{A3}
$$

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

$$
\begin{aligned}
\mathcal{M}(x, r_0, \varrho, r_{\mathrm{cut}}) &= \theta(r_{\mathrm{cut}} - x) \times \\
&\left[\left(e^{\varrho\left(1 - \frac{x}{r_0}\right)} - 2\right)e^{\varrho\left(1 - \frac{x}{r_0}\right)} - V_{\mathrm{shift}}(r_{\mathrm{cut}})\right]
\end{aligned}
$$

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

Cargo-cargo interactions

The interaction between cargo particles is given by

$$
U_{\mathrm{LJ}}(\{\mathbf{l}_i\}, \{\mathbf{l}_j\}) = \sum_{i<j}^{N_l} \varepsilon_{\mathrm{CC}}\mathcal{L}\big(|\mathbf{l}_i - \mathbf{t}_j|,\, \sigma_{\mathrm{C}}, r_{\mathrm{cut}}^c\big)
\tag{A5}
$$

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

$$
\begin{aligned}
\mathcal{L}(x, \sigma, r_{\mathrm{cut}}) &= \theta(x - r_{\mathrm{cut}}) \times \\
&\left\{4\left[\left(\frac{x}{\sigma}\right)^{12} - \left(\frac{x}{\sigma}\right)^6\right] - V_{\mathrm{shift}}(r_{\mathrm{cut}})\right\}
\end{aligned}
\tag{A6}
$$

and $\varepsilon_{\mathrm{CC}}$ is an adjustable parameter which sets the strength of the cargo-cargo interaction, N_l is the number of LJ particles, σ_{C} is the cargo diameter set to $1.0 r_{\mathrm{b}}$ except where mentioned otherwise, and r_{cut}^c is set to $3\sigma_{\mathrm{C}}$.

Subunit-cargo interactions

The subunit-cargo interaction is a short-range repulsion between cargo-excluder and cargo-'Top' pairs reresenting the excluded volume plus an attractive interaction between the cargo - 'Bottom' pairs. For subunit i with excluder positions $\{\mathbf{x}_{i\alpha}\}$ and 'Bottom' psuedoatom $\{\mathbf{b}_{i\alpha}\}$ and cargo particle j with position $\mathbf{R}_j$, the potential is:

$$
\begin{aligned}
U_{\mathrm{Ads}}(\{\mathbf{x}_{i\alpha}\}, \mathbf{R}_j) \ &= \sum_{\alpha}^{N_{\mathrm{x}}} \mathcal{L}\big(|\mathbf{x}_{i\alpha} - \mathbf{R}_j|, \sigma_{\mathrm{ex}}\big) \\
&+ \sum_{\alpha}^{N_{\mathrm{t}}} \mathcal{L}\big(|\mathbf{t}_{i\alpha} - \mathbf{R}_j|, \sigma_{\mathrm{t}}\big) \\
&+ \sum_{\alpha}^{N_{\mathrm{b}}} \varepsilon_{\mathrm{SC}} \mathcal{M}\big(|\mathbf{c}_{i\alpha} - \mathbf{R}_j|, \, r_0, \varrho, r_{\mathrm{cut}}\big)
\end{aligned}
$$

where $\varepsilon_{\mathrm{SC}}$ parameterizes the shell-cargo interaction strength, N_{x}, N_{t}, and N_{b} are the numbers of excluders, 'Top', and 'Bottom' pseudoatoms on a shell subunit, $\sigma_{\mathrm{ex}} = 0.5 r_{\mathrm{b}}$ and $\sigma_{\mathrm{t}} = 0.5 r_{\mathrm{b}}$ are the effective diameters of the Excluder - cargo and 'Top' - cargo repulsions, r_0 is the minimum energy attractor distance, set to $0.5 r_{\mathrm{b}}$, ϱ is a parameter determining the width of the attractive interaction, set to $2.5 r_{\mathrm{b}}$, and r_{cut} is the cutoff distance for the attractor potential set to $3.0 r_{\mathrm{b}}$.

Motivation for choice of interaction potentials

The choices we have made for potential functions (Morse or Lennard-Jones) between different classes of pseudoatoms are based on the need for tunability of the interaction length scale and the extent to which guidance on parameterization is available from the existing literature. In particular, the Morse potential enables controlling the interaction length scale independently from the particle excluded volume size, whereas the interaction length scale and excluded volume size are tuned by a single parameter in the Lennard-Jones potential. Our shell-shell interaction potential is based on previous models for viral capsid assembly (*Wales, 2005*; *Johnston et al., 2010*; *Ruiz-Herrero and Hagan, 2015*; *Perlmutter et al., 2013*; *2014*; *Perlmutter and Hagan, 2015b*), and the choice of a Morse potential for attractor-attractor interactions and a Lennard-Jones potential for Top-Top interactions follows these previous works. The attractor interactions are modeled using a Morse potential because the length scale of their interaction strongly affects the subunit orientational specificity. We chose to model the cargo-cargo interaction using a Lennard-Jones potential because the phase behavior for this model has been extensively studied in the literature, thus limiting the need for model parameterization. However, we note that it could be of interest to study how the probability of shell detachment depends on the length scale of the cargo-cargo interaction; we speculate that a longer-range interaction would increase the probability of detachment by making it easier for shell subunits to penetrate into the globule. Finally, the shell-cargo interactions could have used either choice of potential; we elected to use a Morse potential due to its greater flexibility.

1.2 Maximum cargo loading

To give context to the densities of packaged cargo particles that we observe in simulations, we estimate the maximum possible cargo loading here. Our assembled shell has the geometry of a truncated icosahedron with an edge length of approximately $1.5 d_{\mathrm{u}}$. Accounting for the volume occluded to cargo particles by the shell pseudoatoms, the interior volume is $V_{\mathrm{in}} \approx 109 d_{\mathrm{u}}^3$. The maximum number of cargo molecules that can be packaged (assuming hexagonal close packing) is thus $N_{\mathrm{HCP}} \approx 154$. However, this is an overestimate since the shell geometry is not commensurate with perfect hexagonal close packing. We thus estimate $N_{\mathrm{HCP}} = 148$, the

maximum number of packaged cargo particles seen in an equilibrium simulation. The maximum cargo loading for random close packing is then $N_{\mathrm{RCP}} \approx 120$.

Appendix 2 Thermodynamics of assembly around a fluid cargo

2.1 General theory

In this section we present a general thermodynamic description for assembly around a fluid cargo. The theory provides a description of phase behavior in terms of simple physical parameters, and enables evaluating the extent to which our finite-time dynamical simulations have approached equilibrium. We assume that the equilibrium distribution is dominated by three classes of system configurations: free cargo and shell subunits, a disordered globule of cargo molecules with unassembled shell subunits on its surface, and assembled shells filled with cargo molecules. Extension to consider partially assembled intermediates and partially filled shells is straightforward but would complicate the presentation; moreover, at conditions leading to productive assembly, concentrations of partially assembled intermediates are negligible at equilibrium (*Hagan, 2009*; *2014*; *Safran, 1994*; *Gelbart et al., 1994*).

We consider shells composed of species $\alpha = 1, 2, \ldots M$, with n_α^{shell} subunits of species α in a complete shell, which encapsulates n_0 cargo molecules (the index 0 refers to cargo molecules henceforth). Assembly occurs from a dilute solution of cargo molecules with density ρ_0, shell subunits with density ρ_α for each species, and the density of assembled, full shells as ρ_{shell}. These are in equilibrium with a globule containing n_0^{glob} cargo molecules and n_α^{glob} subunits for each species α. The total free energy density is then given by

$$
\begin{aligned}
f_{\text{tot}} \;=\; & \sum_{\alpha=0}^{M} k_{\text{B}} T \rho_\alpha \left[\ln(\rho_\alpha v_0) - 1 \right] + k_{\text{B}} T \rho_{\text{shell}} \left[\ln(\rho_{\text{shell}} v_0) - 1 \right] \\
& + \rho_{\text{shell}} G_{\text{shell}} + V^{-1} G_{\text{glob}} \left(n_0^{\text{glob}}, \{ n_\alpha^{\text{glob}} \} \right)
\end{aligned}
\tag{B1}
$$

where the sum runs over free cargo molecules and shell subunits, V is the system volume, v_0 is a standard state volume, G_{shell} is the interaction free energy of the assembled shell, and $G_{\text{glob}}(n_{\text{s}}^{\text{glob}}, n_0^{\text{glob}})$ is the globule free energy. We then minimize f_{tot} with respect to $N_{\text{shell}} = V \rho_{\text{shell}}$ and $\{ n_\alpha^{\text{glob}} \}$, subject to the conservation of mass constraints:

$$
\rho_\alpha^{\text{T}} = \rho_\alpha + n_\alpha^{\text{glob}}/V + \rho_{\text{shell}} n_\alpha^{\text{shell}} \quad \text{for } \alpha = 0 \ldots M
\tag{B2}
$$

where ρ_α^{T} denotes the total density of species α.

This results in *Equations (1–2)* of the main text.

2.2 Specification to our computational model

Equations (1–2) are the general equilibrium description for a system of assembling shells with a disordered-phase intermediate. To explore how assembly depends on the control parameters (ε_{CC}, ε_{SC}, ε_{SS}, $\rho_{\text{s}}^{\text{T}}$, and $\rho_{\text{s}}^{\text{T}}$) and to compare these equilibrium expressions against our simulation results, we now specify these relations to our computational model.

2.2.1 Globule and shell interaction free energies

We model the globule as a liquid droplet of Lennard-Jones (LJ) particles, with shell subunits adsorbed to its exterior surface. For simplicity, we treat shell subunit binding to the globule with the Langmuir adsorption model. To simplify the notation, we suppress dependencies on

control parameters in the free energy expressions, but list them beneath. The free energy of the globule is then given by

$$
\begin{aligned}
G_{\text{glob}} \;(n_{\text{p}}^{\text{glob}}, n_{\text{h}}^{\text{glob}}, n_0^{\text{glob}}) = \\
\gamma A_{\text{glob}}(n_0^{\text{glob}}) + \mu_{\text{liq}} n_0^{\text{glob}} \\
+ g_{\text{Ads}}(n_{\text{p}}^{\text{glob}} + n_{\text{h}}^{\text{glob}}) \\
+ G_{\text{mix}}\left(n_{\text{p}}^{\text{glob}}, n_{\text{h}}^{\text{glob}}, n_{\text{max}}(A_{\text{glob}}(n_0^{\text{glob}}))\right),
\end{aligned}
\tag{B3}
$$

where $\gamma(\varepsilon_{\text{CC}})$ and $\mu_{\text{liq}}(\varepsilon_{\text{CC}})$ are the bulk surface tension and chemical potential of a LJ liquid, $g_{\text{Ads}}(\varepsilon_{\text{SC}})$ is the shell subunit absorption free energy, $n_{\text{p}}^{\text{glob}}$ and $n_{\text{h}}^{\text{glob}}$ are the numbers of adsorbed pentamers and hexamers respectively, $A_{\text{glob}} = (\sqrt{3}4\pi\rho_{\text{liq}}(\varepsilon_{\text{CC}})n_0^{\text{glob}})^{2/3}$ is the area of the globule, and $\rho_{\text{liq}}(\varepsilon_{\text{CC}})$ is the density of the LJ liquid. The final term is the mixing entropy of adsorbed subunits according to Langmuir adsorption, given by

$$
\begin{aligned}
G_{\text{mix}}(n_{\text{p}}^{\text{glob}}, n_{\text{h}}^{\text{glob}}, n_{\text{max}})/k_{\text{B}}T = \\
\ln\left(\begin{matrix} n_{\text{max}} \\ n_{\text{p}}^{\text{glob}}, n_{\text{h}}^{\text{glob}}, n_{\text{max}} - (n_{\text{p}}^{\text{glob}} + n_{\text{h}}^{\text{glob}}) \end{matrix}\right),
\end{aligned}
\tag{B4}
$$

with n_{max} as the number of adsorbed subunits at saturation (calculated from simulations, see below).

For the free energy of shell assembly, we consider a shell comprised of $n_{\text{pent}}=12$ pentamers and n_{hex} hexamers, which have n_{ph} pentamer-hexamer contacts with binding energy ε_{ph} and n_{hh} hexamer-hexamer contacts with energy ε_{hh}. For our $T=3$ model, $n_{\text{hex}}=20$, $n_{\text{ph}}=60$, and $n_{\text{hh}}=30$. The assembly free energy is then given by

$$
\begin{aligned}
G_{\text{shell}} = \; n_{\text{ph}}\varepsilon_{\text{ph}} + n_{\text{hh}}\varepsilon_{\text{hh}} \\
- T\left(n_{\text{pent}}s_{\text{pent}} + n_{\text{hex}}s_{\text{hex}} + s_{\text{config}}\right) \\
+ \gamma A_{\text{glob}}(n_0^{\text{glob}}) + \mu_{\text{liq}}n_0^{\text{glob}} + g_{\text{Ads}}(n_{\text{pent}} + n_{\text{hex}}),
\end{aligned}
\tag{B5}
$$

with s_{pent} and s_{hex} the translational and rotational entropy penalty associated with binding of pentameric or hexameric subunits and s_{config} accounting for the configurational entropy associated with subunit and shell symmetries. In our model the pentamers, hexamers, and capsid are 5-fold, 3-fold, and 60-fold symmetric, giving $s_{\text{config}} = k_{\text{B}}\ln(5^{n_{\text{pent}}}3^{n_{\text{hex}}}/60)$. Other parameters were calculated from simulations, as described next.

2.2.2 Determination of parameter values

Since our interactions are constructed from standard potential functions, some of the parameters discussed in the last section are known from the literature, and others can be calculated from simulations. Thus, it is possible to compare our equilibrium theory against simulation results with no fitting parameters. We present the parameter values and how they are obtained in this section.

Cargo parameters

The parameters characterizing the phase behavior of a Leonard-Jones fluid, γ, μ_{liq}, and ρ_{liq} can be obtained from the literature, but we performed fits specific to the parameter ranges of interest, $1.0 \leq \varepsilon_{\text{CC}}/k_{\text{B}}T \leq 3.0$. The surface tension γ was estimated using the approach of Mecke et al. (**Mecke et al., 1997**). We performed separate simulations containing only LJ particles, with numbers of particles and volume for each system set to achieve formation of a

planar liquid vapor interface, and varying values of the LJ interaction strength $\varepsilon_{\mathrm{CC}}$. We then calculated γ from the virial expression. For our LJ potential, truncated at $r_{\mathrm{cut}} = 3\sigma$, we obtain (using the functional form of Ref. (**Mecke et al., 1997**)

$$\gamma(\varepsilon_{\mathrm{CC}}) = 2.936\left(1 - \frac{\varepsilon_{\mathrm{CC}}^{-1}}{1.3}\right)^{1.688}. \tag{B6}$$

From the same simulations, we calculated the dependence of the bulk liquid density on $\varepsilon_{\mathrm{CC}}$ as

$$\rho_{\mathrm{liq}}(\varepsilon_{\mathrm{CC}}) = -1.439 + 2.165\varepsilon_{\mathrm{CC}}^{0.115}. \tag{B7}$$

Although there are a number of empirical forms for the LJ equation of state available in the literature, they vary widely in complexity, number of fit parameters, and presumably accuracy over the parameter range we are interested in. We therefore estimated the liquid chemical potential μ_{liq} from the vapor-phase densities ρ_{vap} in LJ liquid-vapor coexistence simulations according to

$$\mu_{\mathrm{liq}} = k_{\mathrm{B}}T\ln(\rho_{\mathrm{vap}}\sigma^3) - A\gamma/N_{\mathrm{liq}}, \tag{B8}$$

where A is the interfacial area, N_{liq} is the total number of particles in the liquid phase as a function of $\varepsilon_{\mathrm{CC}}$, and γ is given by **Equation B6**. The results are fit well by the linear function

$$\mu_{\mathrm{liq}}(\varepsilon_{\mathrm{CC}}) = 3.13k_{\mathrm{B}}T - 5.6\varepsilon_{\mathrm{CC}}. \tag{B9}$$

Shell subunit-subunit interactions

We estimated the subunit-subunit binding free energy values as functions of the well-depth parameter $\varepsilon_{\mathrm{SS}}$ by measuring the dimerization equilibrium constant in simulations of subunits only capable of forming dimers (**Figure 1C**). For both pentamer-hexamer and hexamer-hexamer dimers, we obtain binding free energies which are linear functions of the well-depth $\varepsilon_{\mathrm{SS}}$. We interpret the y-intercept as the binding entropy, giving:

$$\begin{aligned} g_{\mathrm{ph}} &= \varepsilon_{\mathrm{ph}}\varepsilon_{\mathrm{SS}} - Ts_{\mathrm{pent}} \\ \varepsilon_{\mathrm{ph}} &= -2.95; \quad s_{\mathrm{pent}} = -17.2k_{\mathrm{B}} \\ g_{\mathrm{hh}} &= \varepsilon_{\mathrm{hh}}\varepsilon_{\mathrm{SS}} - Ts_{\mathrm{hex}} \\ \varepsilon_{\mathrm{hh}} &= -3.15; \quad s_{\mathrm{hex}} = -17.7k_{\mathrm{B}} \end{aligned}$$

where the standard state volume is d_{u}^3.

In **Equation (B5)** we then make the assumption that, because the interactions are orientationally specific, a subunit incurs its entire binding entropy penalty upon dimerization — because a bound subunit is already aligned to form additional interactions, these interactions do not lead to further entropy penalties. In reality, this is an under-prediction since some additional entropy losses occur on making additional bonds (**Hagan and Chandler, 2006**; **Hagan et al., 2011**), but these are not sufficiently large to qualitatively affect our results.
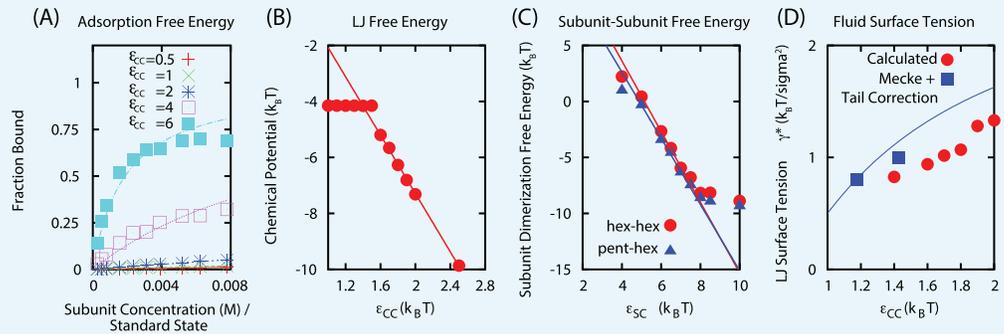
Appendix 2—figure 1. (**A**) Langmuir isotherms to estimate $g_{\mathrm{Ads}}(\varepsilon_{\mathrm{SC}})$. (**B**) Estimate of the chemical potential for an equilibrated LJ system (before correcting for the finite size of liquid droplet). (**C**) Fit of the subunit dimerization free energies $g_{\mathrm{hh}}(\varepsilon_{\mathrm{SS}})$ and $g_{\mathrm{ph}}(\varepsilon_{\mathrm{SS}})$ as a function of the well depth parameter $\varepsilon_{\mathrm{SS}}$. (**D**) Fit of LJ droplet surface tension, including the tail correction. DOI: 10.7554/eLife.14078.027

Shell subunit adsorption onto globule

We estimated the shell subunit adsorption free energy by performing simulations of subunits which cannot assemble ($\varepsilon_{\mathrm{SS}}{=}0$) in the presence of a cargo globule. We then measured the globule size and number of adsorbed subunits as functions of $\varepsilon_{\mathrm{SC}}$. We found the results could be fit using the Langmuir adsorption model, with the adsorption free energy of a single subunit g_{Ads} as a fit parameter for each value of $\varepsilon_{\mathrm{SC}}$. We assumed that the maximum number of adsorbed subunits (the number of lattice sites in the Langmuir model) does not directly depend on $\varepsilon_{\mathrm{SC}}$, and hence fit this parameter globally, obtaining $n_{\mathrm{max}}{=}80$ for a globule with $n_0^{\mathrm{glob}}{=}300$ cargo molecules. In our calculations we assume that n_{max} is proportional to the globule surface area, consistent with observations from simulations. Our fit resulted in a linear relationship between the adsorption energy and free energy over the range of interest:

$$g_{\mathrm{Ads}} = 0.093 k_{\mathrm{B}}T - 1.17 \varepsilon_{\mathrm{SC}}. \tag{B10}$$