Using Cryogenic Electron Tomography (cryoET) to Determine Rubisco Polymerization Constants in α-Carboxysomes

Co-first: Wenxiang Cao[1], Kristy Rochon[2]

Middle authors: Ryan H. Gray[3], Luke Oltrogge[4,5], David F. Savage[4,5]

Co-corresponding: Enrique M. De La Cruz[1], and Lauren Ann Metskas[2,3]

1 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, United States

2 Department of Biological Sciences, Purdue University, West Lafayette, IN, United States

3 James Tarpo Jr. and Margaret Tarpo Department of Chemistry, Purdue University, West Lafayette, IN, United States

4 Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, United States

5 Howard Hughes Medical Institute, University of California; Berkeley, CA, United States

Classification: Biological Sciences, Biophysics and Computational Biology
Key words: polymerization, cryogenic electron tomography, bacterial microcompartments, carboxysomes, Rubisco

Abstract

Bacteria microcompartments (BMCs) are pseudo-organelles comprised of a self-assembling, semi-permeable protein shell, most commonly enclosing components of enzymatic pathways. α-Carboxysomes (α-CBs) are anabolic BMCs known for their role in sequestering Rubisco, the enzyme responsible for carbon fixation in plants, algae and bacteria, along with an upstream enzyme and an assembly protein. Rubisco has low selectivity for its substrate, CO_2, and has a slow enzymatic turnover rate, resulting in an inefficient metabolic pathway. Within the α-CB, Rubisco has been observed at a range of concentrations and with either a liquid-like assembly or a pseudo-lattice of polymerized fibrils. The biophysical origins of the fibril ultrastructure organization are unclear; however, it is only observed inside α-CBs. Quantitative knowledge of the binding constants and energies for assembly and maintenance of these fibrils is critical for understanding this organization and Rubisco regulation, but quantitative methods for *in situ* analysis of Rubisco polymerization have been lacking. Here, we present an approach to convert tomography-derived α-CB volumes and Rubisco particle positions into polymerization binding curves. We used this procedure to determine the Rubisco polymerization constants, including the nucleus size (n) and equilibrium polymerization constant (K_{pol}). The adopted modeling approach is consistent with *in situ* constraints, such as concentration-dependent binding interactions and confinement. This approach offers a powerful tool to evaluate both *in vitro* and potentially *in vivo* biomolecular interactions, both of Rubisco and of other proteins and polymers suitable for analysis by cryo-electron tomography.

Significance Statement

Cryogenic electron tomography (cryoET) is a powerful method to resolve structures of proteins in their native environment at subnanometer-level resolution. Because tomography data retains spatial relationships of all particles, it intrinsically contains information about component (e.g., protein) binding interactions. Here, we use Rubisco polymerization in α-carboxysomes as a model system to demonstrate that quantitative, biochemical binding analysis is possible with cryoET.

Introduction

Quantitative knowledge of the equilibrium binding affinities of (macro) molecular interactions is crucial for understanding the regulation of key biological processes including enzyme catalysis, signal transduction, gene expression, and the assembly of macromolecular complexes (1–4). Established methods for determining molecular binding constants commonly rely on experimental data acquired *in vitro*, under controlled experimental conditions (5, 6). Such *in vitro* studies are indispensable for developing quantitative, mechanistic models, but they often have inherent limitations, such as restricted concentration ranges, non-native chemical modifications (i.e., tags), and/or minimal buffer systems substituting for uncharacterized physiological chemical environments. Studies rarely evaluate the effects of physiological crowding (7, 8), and thus fail to replicate native physiological conditions.

Cryogenic electron tomography (cryoET) is a powerful imaging technique with resolutions ranging from near-atomic (protein structures) to micron (cellular features) (9–11). While the method typically provides slightly lower resolution than cryogenic electron microscopy (cryoEM) in structural biology applications, it has the unique ability to preserve the spatial relationships between imaged particles. In subtomogram averaging (STA), proteins are identified, extracted from the larger image, aligned and averaged, and then replaced into the tomogram, enabling extrapolation of these spatial relationships (12, 13). Despite these capabilities, the most detailed analyses of proteins with cryoET typically focus on structure, with more qualitative or morphological descriptions of protein environment, proximity to other features, and ultrastructure arrangement.

Here, we present a method using cryoET to quantify binding interactions and polymerization constants of a self-associating biopolymer *in situ*. We posit that if all particles in a defined and precisely measured volume are accurately identified and their oligomeric states distinguished, biophysical binding parameters such as polymerization constants can be calculated from STA data. In addition to providing simultaneous insight into protein conformation and binding, this approach eliminates the need for complex *in vitro* assay design while maximizing the data available in tomograms and extending polymerization assay capabilities to systems not amenable to *in vitro* measurements.

We use Rubisco in α-carboxysomes (α-CBs), a bacterial microcompartment (BMC) responsible for carbon fixation in *Halothiobacillus neapolitanus* (14–16), as a model system to validate this approach.

α-CBs are ~1 aL pseudo-organelles with a selectively permeable protein shell (17–20). The protein shell assembles around an enzyme condensate, facilitated by a disordered scaffolding protein (21). α-CBs encapsulate Rubisco with its upstream enzyme carbonic anhydrase, maximizing Rubisco turnover rates presumably through substrate saturation (14, 19, 22–24). α-CB assembly and cargo sequestration rely largely on Rubisco condensation through liquid-liquid phase separation (25). This process requires a disordered accessory protein (CsoS2) interfacing with the shell to assemble the facets while also forming interfaces with Rubisco to facilitate inclusion within the microcompartment (21). The shell is assembled around the condensate.

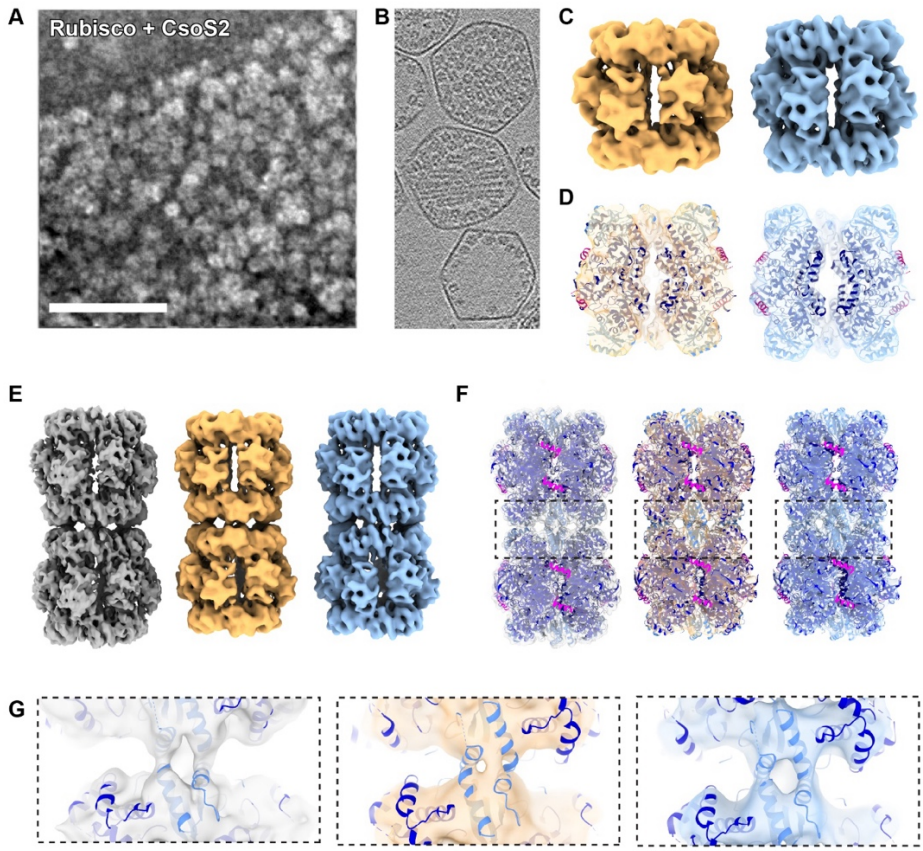


Figure 1. Aligned Rubisco Particles from Tomograms. A. Negative stain TEM micrograph of a Rubisco-CsoS2 condensate. **B.** Orthoslice of a tomogram containing α-CBs with examples of Rubisco packing: dense (top), ordered (middle), and sparse (bottom). **C.** Rubisco densities resulting from STA of DTT treated (left, orange) and diamide treated (right, blue) samples. **D.** Docked crystal structure (PBD ID: 6UEW) in DTT-treated density (left, yellow) and diamide-treated density (right, blue). **E-G.** Composite chains of STA results with docked crystal structures of (PDB ID: 6UEW) for untreated Rubisco (left, gray), DTT-treated (middle, orange) and diamide-treated (right, blue). **G.** Isolated Rubisco-Rubisco interfaces resolved using STA. Untreated (left, gray), DTT-treated (middle, orange) and diamide-treated (right, blue).

Rubisco condensates with CsoS2 inside α-CBs assemble pseudo-lattices of polymer fibrils, a behavior not replicated with purified components *in vitro* (Fig. 1A) (21, 26, 27). Recent studies have indicated that BMC shells can separate redox states and that the interior of β-CBs becomes oxidized over time, with unknown implications on the BMC's interior organization or function (28, 29).

In this study, we developed and validated a framework and procedure to quantitatively extract polymerization binding constants directly from cryoET data, using Rubisco in α-CBs as a model system. By converting STA volumes and particle positions into polymerization “binding” curves, we determined key binding equilibrium constants governing Rubisco fibril assembly *in situ*. This approach not only addresses a critical gap in our ability to quantify biomolecular assembly under native physiological conditions, offers a complementary technique to traditional *in vitro* methods and extending the analytical applications of cryoET beyond structural studies, but also provides a general method in analysis of biomolecule polymerization with the theories already developed.

Results

Rubisco oligomeric states within α-CBs

Rubisco within α-CBs can adopt two oligomeric/polymer states (Fig. 1B): dense (disorganized or liquid-like) and ordered (lattice-like packing). Purified α-CBs can also have low Rubisco concentration (termed "sparse") as an experimental consequence of centrifugation (26). It is not known what determines whether Rubisco within α-CBs adopts the dense or ordered morphology at equivalent concentrations. We hypothesize that Rubisco polymerization could be affected by redox state. To test this, α-CBs were isolated from *H. neapolitanus* and treated with either DTT (reducing agent) or diamide (oxidizing agent) to alter their internal redox state before vitrification and cryoET analysis of polymerization.

Rubisco monomers, oligomers, and linear polymers were identified within intact α-CBs and STA performed out to sub-7Å resolution using published methods (Fig. 1C) (26, 30). Helices are well defined at this resolution, allowing precise estimation of position and orientation for each Rubisco complex (Fig. 1D). We used a numerical classification approach to calculate α-CB volume and determine the binding state of each Rubisco complex, using the positions and orientations generated by STA analysis (31). These individual Rubisco are classified as “bound” (oligomer or polymer; nuclei are treated as polymers; this is necessary for mass balance in the theory) or “unbound” (monomer for nucleus size n = 2, or monomer plus dimer for nucleus size n = 3; discussed below). The bound species can be categorized by the number (n) of contiguous subunits comprising the polymer. A Rubisco identified as part of a polymer chain can be further processed with STA to isolate the Rubisco-Rubisco subunit interface (Fig. 1E-G). By using a numerical analysis rather than image classification, the analyst controls the parameters and corresponding error rates and can cleanly establish the polymer length and connectivity of each Rubisco; it is also computationally more efficient.

Rubisco packing within α-CBs

The average volume of an untreated α-CB is $7 \pm 2 \times 10^{-4}$ μm^3 corresponding to approximately a cube with 90 nm edge length or a sphere with 60 nm radius (Fig. 2A, Table 1). DTT treatment changed the average volume (Table 1), possibly through destabilization of the shell leading to selective loss of larger α-CBs (Fig. 2A, Table 1). Both chemical treatments resulted in populations with slightly higher

Rubisco concentrations (45% for DTT treatment, 26% for diamide treatment, Table 1), presumably due to loss of less mechanically stable α-CBs during plunge-freezing. While the chemical treatments reduced the number of low-concentration data points available for polymerization analysis, enough remained to facilitate curve fits.

Table 1. Rubisco packing properties inside α-CB.

	untreated	DTT-treated	Diamide-treated	Data
Packing (average values)				
α-CB Volume(10^{-4} μm^3)	6.9 ± 2.1	5.3 ± 1.4	8.5 ± 3.4	Fig. 2A
Rubisco complexes per α-CB	242 ± 73	260 ± 77	363.4 ± 160	Fig. 2B
Rubisco concentration (μM)	583 ± 106	822 ± 100	718 ± 139	Fig. 2C
Rubisco complexes per 10^{-3} μm^3	340 ± 6	494 ± 7	427 ± 7	Fig. 2D
Rubisco [inner] vs. [total] slope	1.48 ± 0.07	1.30 ± 0.06	1.38 ± 0.04	Fig. S2
[inner] vs. [total] *x*-intercept (μM)	148 ± 5	82 ± 22	68 ± 6	Fig. S2

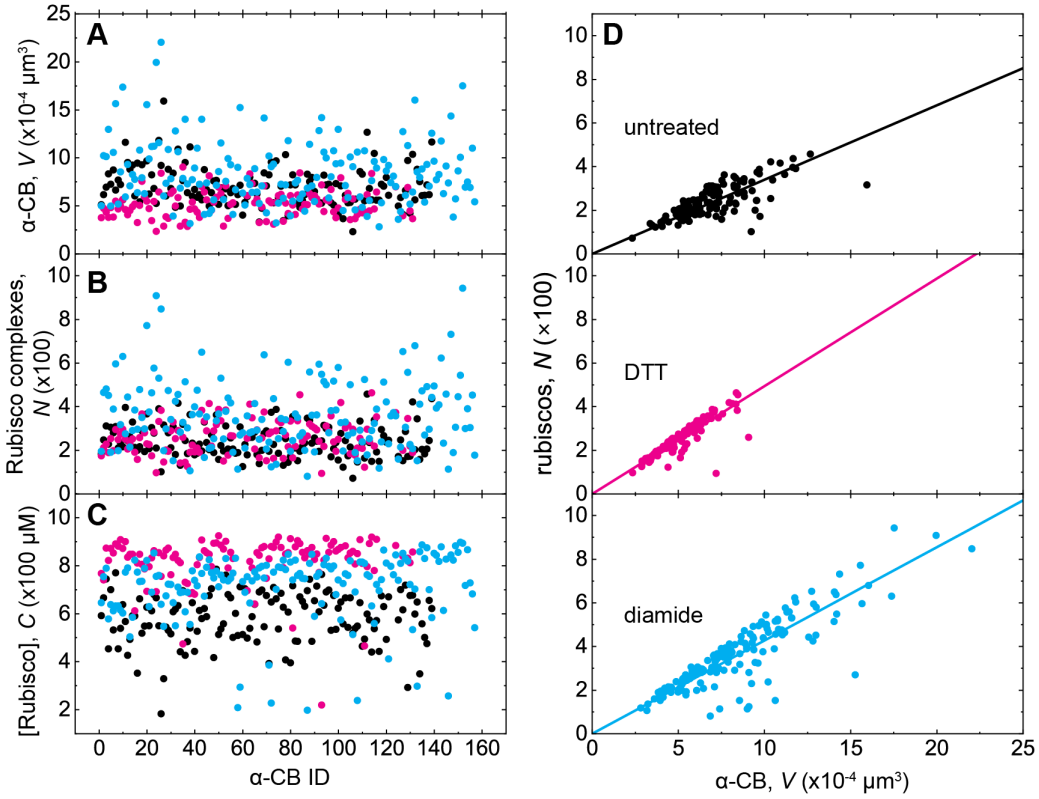


Figure 2. Rubisco packing in α-CBs. A-C. Comparisons of α-CB volume (**A**), number of Rubisco complexes in α-CB (**B**) and Rubisco concentration (A_{tot}) in α-CBs (**C**) of untreated (black), DTT-treated (magenta), and diamide-treated (blue) samples. Data are summarized in Table 1. **D.** Correlation between the number of Rubisco complexes packed inside an α-CB vs. the volume.

The concentration of inner Rubiscos (defined as those not adjacent to the α-CB shell) scaled linearly with the total Rubisco concentration in untreated, DTT-treated, and diamide-treated samples (Fig. S2A-C). On average, inner Rubisco concentrations are about 1.4-fold higher than their total Rubisco concentrations for all three treated and untreated samples. The best fit lines to the inner Rubisco concentrations to total Rubisco concentrations for all three α-CB samples have a positive *x*-axis intercept (Fig. S2A-C, Table 1). This behavior indicates that inner Rubiscos are absent and all Rubisco is contained within the shell region at low total Rubisco concentrations, consistent with tomograms of lower-concentration α-CBs. This suggests that Rubisco near the shell may have different properties, interactions or other behaviors compared to the α-CB interior.

Rubisco polymer chain length distribution within α-carboxysomes.

Rubisco forms polymer chains in all α-CB populations imaged (Fig. 1B). Most of the polymers observed in the data are dimers, and the concentrations of longer chains decay with the chain length (Fig. 3). The Rubisco in dimers observed are not randomly contacting each other; rather, the interfaces observed between adjacent, continuous subunits in the Rubisco dimers are the same as those found in the longer polymers (Experimental section).

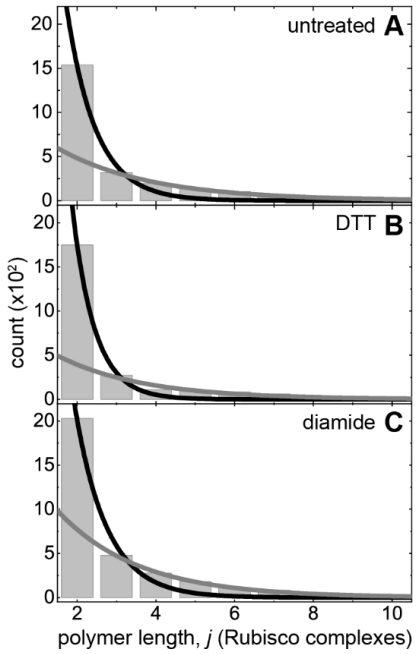


Figure 3. Length distribution of Rubisco fibrils in α-CBs. A-C. Histogram of polymer length for untreated, DTT-treated or diamide-treated. The continuous black lines are the best fits of the data to exponential decays including the dimer population (the first bar). The continuous gray lines (starting lower) are the best fits excluding dimers. Data available in Table 2.

The length distribution of an equilibrium polymer follows an exponential decay (32) (see Eq. 6 in Theory section). Histograms of Rubisco chain lengths inside all untreated, DTT treated and diamide treated α-CBs were fitted to an exponential (Fig. 3). The best fit to all the polymer lengths including dimers ($j \geq 2$ Rubisco complexes; black smooth line) does not fit the data well, especially at longer polymer lengths. The observed dimer population deviates from the exponential decay defining the polymer population distribution, indicating that dimers are thermodynamically distinct from polymers of three or more Rubisco complexes. In contrast, the best fit to the polymer length distribution for trimers and longer polymers ($j \geq 3$ Rubisco complexes; gray smooth line) fits the data well for all samples (Fig. 3). This exponential length distribution behavior indicates (Eq. 6) that the internal Rubisco interactions in polymers of three or more Rubisco complexes ($j \geq 3$) are thermodynamically equivalent. We interpret this as an indication that the nucleus for Rubisco polymerization is three contiguous Rubisco complexes (n = 3).

A previous thermodynamic theory analysis (33) has shown that the chain length exponential decay constant (κ) is directly proportional to the total free energy of the polymer ($G_j^{o\prime}$ for a polymer composed of j Rubisco complexes) and the free energy change of the polymer ($\Delta G_{j\to j+1}^{o\prime}$) resulting from elongation by one subunit, in turn, can be linked to the polymerization constant (K_{pol}; also referred to as the "critical concentration" for polymerization (34)) and the free monomer concentration in equilibrium with polymer (A_1; Eq. 6 in the Theory section) according to:

$$\frac{\Delta G_{j\to j+1}^{o\prime}}{k_B T} = \kappa = -\ln\left(\frac{A_1}{K_{pol}}\right)$$

The free energy change ($\Delta G_{j\to j+1}^{o\prime}$) of the polymer when elongated by one subunit is given by:

$$\Delta G_{j\to j+1}^{o\prime} = \kappa k_B T = k_B T\big(\ln(K_{pol}) - \ln(A_1)\big) \qquad \text{Eq. 10}$$

The free energy change per subunit extension ($\Delta G_{j\to j+1}^{o\prime}$) is constant for given K_{pol} and A_{tot}, so the total value of $\Delta G_{j\to j+i}^{o\prime} = i\kappa k_B T$ scales with the total number (i) of subunits incorporated into the polymer, including those resulting from the annealing of existing polymers.

The best fit to the Rubisco polymer ($j \geq 3$) length distribution to an exponential (Fig. 3 and Table 2) yields $\Delta G^{o\prime}_{j\rightarrow j+1}$= +0.42 ± 0.03, +0.46 ± 0.08, and +0.49 ± 0.02 k_BT for untreated, DTT treated and diamide treated Rubisco complexes, indicating the Rubisco polymer chain internal stability is < k_BT and comparable for all three samples. We emphasize that this free energy is distinct from the standard Gibbs free energy change ($\Delta G^{0\prime}{}_{pol}$) associated with subunit addition to the filament end (discussed below).

Table 2. Rubisco fibril chain properties inside α-CB.

	untreated	DTT-treated	Diamide-treated	Data
Distance between adjacent fibrils (nm)	10.9 ± 0.7	10.3 ± 0.5	10.4 ± 0.6	cryoET
Maximum fibril chain length, *m* (Rubisco)	10	10	11	Fig. 3
length decay const., κ (+dimer; Rubisco^{-1})	1.3 ± 0.2	1.8 ± 0.2	1.2 ± 0.1	Fig. 3
length decay const., κ (–dimer; Rubisco^{-1})	0.42 ± 0.03	0.46 ± 0.08	0.49 ± 0.02	Fig. 3

Rubisco polymerization mechanism within α-CBs

Because we determined the nucleus for Rubisco polymerization is a trimer (*n* = 3; Fig. 3), we treat Rubisco trimers and longer polymers as "bound" when plotting the concentration-dependence of polymer concentration (blue filled circles in Figs. 4 and S4 right panels). In this case, Rubisco dimers belong to the unbound, 'monomer pool' (light blue filled circles in Figs. 4 and S4 right panels). To extract the polymerization binding constants from this data, we can either fit the A_{tot} -dependence of bound Rubisco alone or globally fit both unbound and bound Rubisco concentrations together. Both fitting methods yield similar results, so we present only the global fits.

The best global fit of the data to a length unlimited nucleation model (Eq. 3 and 4) with a trimer nucleus (*n* = 3) yields a K_{pol} = 541 ± 44 μM for untreated Rubisco, 762 ± 5 μM for DTT-treated Rubisco and 730 ± 170 μM for diamide-treated Rubisco polymers. These polymerization constants are weak (K_{pol} >500 μM), and associated with small standard Gibbs free energy changes for incorporation of a Rubisco subunit at a polymer end ($\Delta G^{0\prime}{}_{pol}$) according to:

$$\Delta G^{0\prime}{}_{pol} = k_BT \ln K_{pol} > k_BT \ln(5 \times 10^{-4}\ \mathrm{M}) = -7.6 k_BT \qquad \text{Eq. 11}$$

The nucleation binding constants (K_{nuc}) are even weaker in all three cases, with values ranging from 10^9 to 10^{21} μM (Fig. 4, Table 3).

Table 3. Rubisco polymerization properties inside α-CB.				
	untreated	DTT-treated	Diamide-treated	Data
data: monomer vs. polymer with dimer				
Nucleation Polymerization (from fits), nucleation n = 2,				
K_{pol} (μM, un-limit)*	473 ± 16	654 ± 5	660 ± 13	Fig. 4
K_{nuc} (μM, un-limit)*	9×10^{17}	5×10^{20}	6×10^{6}	Fig. 4
K_{pol} (μM, limit)*	516 ± 10	678 ± 12	686 ± 10	Fig. S4
K_{nuc} (μM, limit)*	5×10^{4}	1×10^{5}	1×10^{5}	Fig. S4
Isodesmic Polymerization (doesn't fit, for comparison only, values from fits)				
K_{pol} (μM, un-limit)*	3705 ± 176	5433 ± 230	7405 ± 324	Fig. S4
K_{pol} (μM, limit)*	3705 ± 208	5433 ± 230	7404 ± 324	Fig. S4
data: monomer+dimer vs. polymer without dimer				
Nucleation Polymerization (from fits), nucleation n = 3,				
K_{pol} (μM, un-limit)*	541 ± 44	762 ± 5	730 ± 170	Fig. 4
K_{nuc} (μM, un-limit)*	6×10^{12}	2×10^{21}	9×10^{9}	Fig. 4
K_{pol} (μM, limit)*	593 ± 15	794 ± 21	762 ± 15	Fig. S4
K_{nuc} (μM, limit)*	6×10^{7}	2×10^{8}	2×10^{8}	Fig. S4

* un-limit: un-limited length model. limit: limited length model.

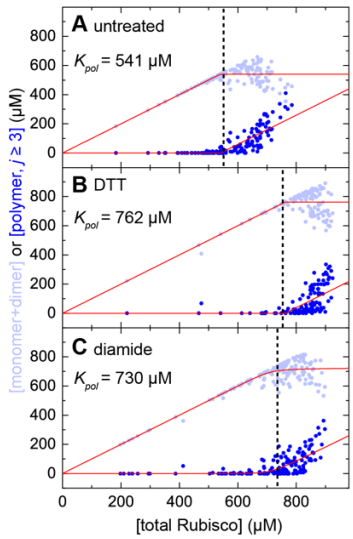


Figure 4. Rubisco polymerization analysis. A-C: Unbound (monomer+dimer, light blue filled circles) and bound (polymer of length $j \geq 3$ Rubisco complexes; blue filled circles) Rubisco concentrations change as a function of the total Rubisco concentration inside untreated, DTT-treated, and diamide-treated α-CB samples. The smooth red lines through the corresponding data points represent the best global fits with implicit nucleation-polymerization equations for unbound (Eq. 3) and bound (Eq. 4) and the nucleation size of n = 3 held, respectively. The dashed vertical lines in all panels indicate the positions of K_{pol}.

For comparison, we analyzed the data with five other polymerization models: length-limited and length-unlimited isodesmic polymerization, length-limited and length-unlimited nucleated polymerization with a dimer nucleus (n = 2; Fig. S4A-C, left), and length-limited nucleated polymerization with a trimer nucleus (n = 3; Fig. S4D-F, right), together with length-unlimited n = 3 nucleated polymerization (Figs. 4 and S4D-F). In the case of isodesmic polymerization and dimer nucleation polymerization, unbound Rubisco complexes are treated as "monomers" (light gray filled circles), and bound Rubisco complexes are dimers and longer polymers (gray filled circles in Fig. S4A-C, left).

Isodesmic polymerization models do not fit the experimental data well for either length-limited or length-unlimited models (Fig. S4A-C, left). Dimer and trimer nucleation models with unlimited polymerization fit the corresponding data equally well (Fig. S4) and considerably better than with a length-limited constraint. The results of these analyses are summarized in Table 3. Fitting such polymerization data cannot determine the size of the nucleus (n) because the nucleus size was constrained when classifying oligomers as bound or unbound, thus, relating fit quality to nucleus size would be circular reasoning.

The best fits of the data to length-unlimited nucleated (n = 2 or 3) polymerization models yield very weak nucleation constant (K_{nuc}) values with large uncertainties for all Rubisco samples (Table 3). The K_{nuc} values also vary by many orders of magnitude among the three samples. Despite the large uncertainties, the K_{nuc} values reflect and capture the sharpness of the monomer-polymer transition (i.e., inflection point) at the macromolecule critical concentration (Fig. S5). Abrupt transitions at the critical concentration indicate that nucleation (K_{nuc}) is much weaker than subunit incorporation at a filament end (K_{pol}). Self-assembling polymer systems with $K_{pol} << K_{nuc}$ and abrupt transitions (i.e., actin filaments; (35)) are said to be "highly cooperative". Thus, Rubisco polymerization is highly cooperative, despite the observed weak critical concentration. We note that despite the weak polymerization and large uncertainties in the collected data, the best fits can readily distinguish changes in the critical concentration for assembly under different conditions (Figs. 4 and S4).

Discussion

A novel method for evaluating protein-protein binding in STA

We demonstrate a novel approach to determine protein binding using volumetric imaging with STA, a technique typically used for protein structure determination (36). This approach leverages the intrinsic capability of STA to provide position and orientation information about proteins and other biological macromolecules, using these data to determine the relative concentrations of bound (i.e., polymerized, in a complex, etc.) and unbound particles in their biological context. This bottom-up method can be implemented as long as a protein can be clearly identified in both bound and unbound states for accurate classification. It is agnostic to concentration, allows experiments under native conditions, and does not require tagging or chemical modification of the target molecules. Concentrations can be calculated either from an enclosing barrier or directly from the particles' occupied volume, either counting the particles alone or including molecular crowding as nonspecific electron density within the particle-occupied area (31).

Single-particle cryoEM has previously been used to estimate binding constants of purified proteins by resolving the structures of bound and unbound states. In structure-based drug design, cryoEM structures are used to select computational binding poses that correlate with compound potency, with binding energies validated experimentally by surface plasmon resonance or estimated computationally through energy minimization of predicted conformational changes upon substrate binding (37, 38). While this approach has shown promise for drug discovery, physiological relevance remains a concern: during plunge-freezing preparations, proteins are exposed to fluid shearing force, uncontrolled osmolarity increases, and surface interactions. Furthermore, the high-resolution structures required for atomic-level comparison preclude inclusion of the entire dataset, as less populated or heterogeneous states would not resolve. CryoET mitigates some of these concerns by imaging proteins within an enclosed system, such as a cell, microcompartment, or artificial lipid vesicle enclosing purified protein. These environments shield the interior from nonphysiological conditions during sample preparation, while supporting more complex three-dimensional analyses of protein organization and interactions.

Our model system, Rubisco in α-CBs, offers several advantages for the development of the method used here to identify protein assembly and determine binding and polymerization constants.

Rubisco complexes are large and easily identifiable on a grid, and the presence of a visible shell with a defined volume provides secondary validation of occupied volume calculations for determining Rubisco concentration (31). The protein shell sequesters the interior chemically and mechanically over the sample preparation timescale, preserving the native conditions. Rubisco's large size allows unambiguous particle identification in bound and unbound forms, and its marked conformational stability allows STA of the full population without the need for classification (30). Finally, the α-CB is relatively well characterized biochemically and relative abundances of the contained proteins known, providing validation for underlying calculations such as Rubisco concentration in the interior (39). Application of this new binding analysis methodology to other systems is certainly possible, with noise and error rates ultimately dependent upon the particle identification accuracy and final resolution of the STA dataset.

Rubisco polymerization in α-carboxysomes

All the three Rubisco polymerization datasets yield an weak (large in value) polymerization constant K_{pol}, (i.e., overall dissociation equilibrium constant for a monomer to bind either end of a polymer chain (40)), and an even weaker nucleation constant K_{nuc} (Table 3), consistent with the observed sharp transitions near the inflection point for all three samples (Fig. 4, S4 and S5). This supports a mechanism in which space confinement is a primary factor for Rubisco polymerization.

We tested four general models of macromolecule polymerization: isodesmic and nucleation models, both with and without spatial constraints. In unlimited space, the length of a self-assembling polymer chain is unconstrained and, in theory, can reach infinite lengths as the total subunit concentration approaches infinity. Under confinement, the polymer chain length is limited by the accessible space, geometry of the polymer chain, and the chain's material properties.

The average diameter of an untreated α-CB is equivalent to the linear length of 10 or 11 Rubiscos complexes (see *Polymerization theory within a confined volume* in Theory of Protein Polymerization), suggesting the volume may limit the maximum allowed Rubisco chain length. The polymerization models must be modified to reflect this spatial constraint (Eq. 7). The compliance of the polymer can also contribute to the achievable length. For example, if Rubisco polymers are very compliant, such that they could bend and twist easily, the maximum allowed chain length could conceivably be longer than a rigid,

linear polymer. In our data, the maximum Rubisco chain length observed is equivalent to the α-CB diameter (10 Rubisco complexes for untreated and DTT-treated and 11 for diamide-treated (Table 2)). This suggests that Rubisco polymers are rigid on the observed length scales and the chain length is limited by the accessible space within α-CBs.

All polymerization models tested here predict that the polymer chain length distribution at equilibrium follows an exponential decay (Eq. 6). The observed Rubisco chain length distribution shows that the dimer fractions deviate from the exponential distribution of the population, whereas trimers and longer polymers adhere to this exponential population distribution (Fig. 3). This suggests Rubisco polymerization follows a nucleation mechanism with nucleus size (n) of 3 Rubisco complexes.

Isodesmic polymerization models, in which nuclei are not required for polymer formation, do not fit data well (Fig. S4A-C). Nucleation models with either $n = 2$ or $n = 3$ with confinement yield reasonable fits to each corresponding data set, although the fits are worse than without confinement (Figs. 4 and S4) as indicated from the sharpness of the transition from unbound to bound species (i.e., the critical concentration inflection point).

The polymerization data were better fitted by the chain-length unlimited nucleation model (Figs. 4 and S4), suggesting that spatial confinement has limited effects on Rubisco polymerization. If the length limit was achieved before subunit packing, the observed filament length distribution would shift toward a uniform population with the spatially-limited length. Therefore, the Rubisco filament length distribution and polymerization analyses together favor a length-unlimited nucleation mechanism with a trimer (n =3) nucleus for Rubisco polymerization in α-CBs.

Previous α-CB assembly studies have returned conflicting results a tomography study observed α-CB-sized Rubisco condensates in the cytoplasm that the authors interpreted as assembly intermediates (41), suggesting that the Rubisco condensate would determine the α-CB size, while biochemical experiments indicate that the accessory protein CsoS2 is the primary determinant of α-CB shell diameter (42–44). We conclude that, while we cannot determine the details of α-CB assembly from these polymerization fits, we can successfully determine that the maximal Rubisco chain length is an intrinsic property of the condensate rather than imposed by the shell.

Disproportional abundance of Rubisco dimers

Rubisco dimers are prevalent in condensates (Fig. 1A) as well as in α-CBs *in vitro* and *in situ*. Conversely, Rubisco trimers and longer polymers are found in low concentrations in solution, and are primarily in α-CBs(21).

We observed a large fraction of Rubisco dimers in α-CBs. This abundance of dimers is unexpected for a nucleus size (n) of 2 or 3, as oligomers shorter than the nucleus (n = 3) are typically unstable and populated at low concentrations, and oligomers equal to the nucleus should behave as polymers in the exponential length distribution (Fig. 3). The deviation of the dimer population from the exponential, polymer length distribution (Fig. 3), suggests that they are thermodynamically distinct from polymerized Rubisco fibrils.

The total dimer species is overpopulated, implying that it is more stable than any polymer species, and raising the possibility that a fraction of this dimer population is polymerization-incompetent. At least four cases can potentially account for this unexpected finding:

Case 1: *Ambiguity identifying and quantifying the Rubisco dimer population*: Two Rubisco monomers positioned in an adjacent manner could be incorrectly classified as a dimer, yielding an "overcounting" of the dimer population. (This would yield n=3 nucleus size, but with some dimers being monomers; see Figs. 4 and S4D-F.) However, the cryoET method employed here retains sufficient spatial and structural resolution to distinguish and confirm that the observed dimers are not randomly contacting each other but adopt the same orientation as adjacent, contiguous subunits in longer polymers. Furthermore, this packing effect should lead to similar "overcounting" of all observed oligomer and polymer species detected. We thus can eliminate miscounting as the source of the observed abundant dimer population.

Case 2. A population of Rubisco dimers are polymerization-incompetent and on the polymerization pathway.

A mechanism in which two distinct Rubisco dimer conformers – one capable of polymerizing (R_2) and one not (R_2*) – are linked by a reversible and on-pathway isomerization equilibrium (K_{isom}) can account for the observed overabundance of Rubisco dimers (Scheme 3a).

$$\overbrace{R + R \underset{}{\overset{K_{nuc1}}{\rightleftharpoons}} R_2^* \overset{K_{isom}}{\rightleftharpoons} R_2 + R \overset{K_{nuc2}}{\rightleftharpoons}}^{K_{nuc} = K_{nuc1}\, K_{isom}\, K_{nuc2}} R_3 + R \overset{K_{pol}}{\rightleftharpoons} R_4 + R \overset{K_{pol}}{\rightleftharpoons} \ldots\ldots$$

Scheme 3a

Given the resolution and orientation criteria for dimer classification (discussed in *Case 1*), the structural difference between the two dimer conformers would need to have a negligible effect on the Rubisco backbone, and likely be located at the subunit interface.

This polymerization pathway (Scheme 3a) defines a Rubisco trimer as the nucleus for polymerization (n = 3). The nucleation constant (K_{nuc}) is defined here in a step-wise manner with binding constants K_{nuc1}, K_{isom}, and K_{nuc2}, such that the overall nucleation equilibrium constant is given by the product of these three step-wise binding constants ($K_{nuc} = K_{nuc1}K_{isom}K_{nuc2}$). This allows both the polymerization-incompetent dimers (R_2*) and the nucleus (R_3) to be expressed in terms of the polymerization-competent dimer (R_2) concentration at equilibrium and the corresponding dissociation equilibrium constants (K_{isom} and K_{nuc2}) according to $[R_2{}^*] = [R_2]K_{isom}$ and $[R_3] = [R_2][R]/K_{nuc2}$.

Accordingly, the ratio of the total dimer concentration (i.e., $[R_{2,tot}] = [R_2^*] + [R_2]$) over the nucleus trimer (R_3) concentration is given by (Eq. 12):

$$\frac{[R_2{}^*] + [R_2]}{[R_3]} = \frac{(1 + K_{isom})K_{nuc2}}{[R]} \qquad \text{Eq. 12}$$

This relationship indicates the observed dimer population ($R_{2,tot}$) will exceed that of the nucleus trimer (R_3) when $(1 + K_{isom})K_{nuc2} > [R] \sim 500\ \mu M$ (i.e., the Rubisco monomer concentration in equilibrium with polymer). Therefore, a weak K_{isom} (i.e., equilibrium largely favors polymerization-incompetent R_2*) and/or weak K_{nuc2} (i.e., equilibrium favors R_2 over R_3) would generate an overpopulated dimer fraction.

In Scheme 3a, subunit addition to a dimer end (K_{nuc2}) could be weaker or tighter than addition to a trimer end (K_{pol}). If the ends of a Rubisco dimer and trimer are identical, as may be expected for a 1-D polymer, K_{nuc2} would be defined as K_{pol}, K_{nuc} would be defined as $K_{nuc1}K_{isom}$, and the nucleus size would be a dimer (R_2, n = 2), as detailed by (Scheme 3b):

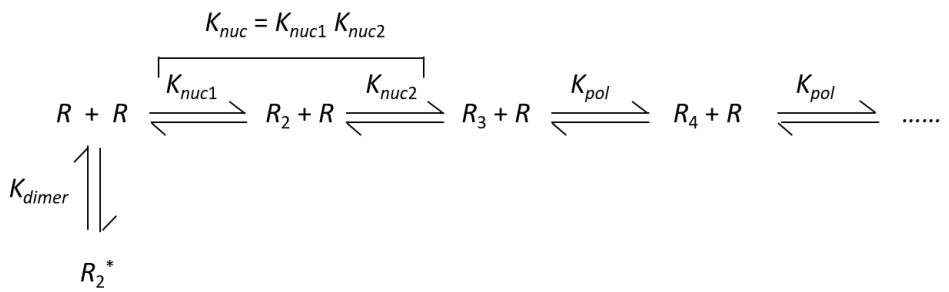
$$\overbrace{R + R \underset{}{\overset{K_{nuc1}}{\rightleftharpoons}} R_2^* \overset{K_{isom}}{\rightleftharpoons}}^{K_{nuc} = K_{nuc1}\, K_{isom}} R_2 + R \overset{K_{pol}}{\rightleftharpoons} R_3 + R \overset{K_{pol}}{\rightleftharpoons} R_4 + R \overset{K_{pol}}{\rightleftharpoons} \ldots\ldots$$

Scheme 3b

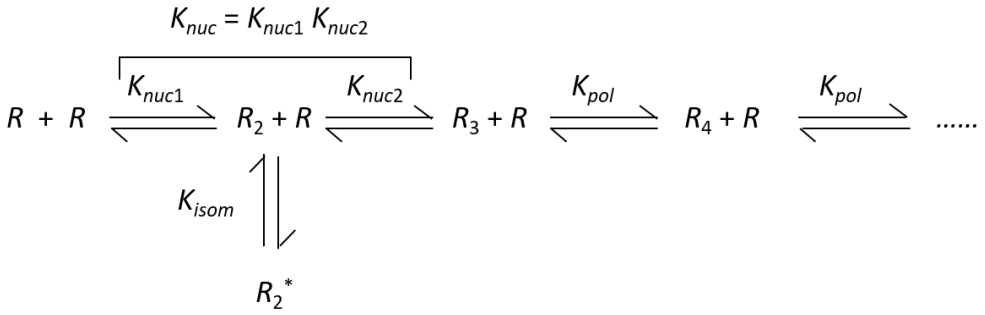
In this case, the nucleus R_2 population adheres to the exponential, polymer-length distribution (Fig. 3), and the polymerization incompetent dimer (R_2^*) contributes to the over-counted dimer fraction. As with Scheme 3a, Eq. 12 with K_{nuc2} defined as K_{pol} predicts that the total observed dimer ($R_{2,tot}$) will be overpopulated when K_{isom} is weak and largely favors R_2^* such that $(1 + K_{isom}) > 1$.

Case 3. *<u>A population of Rubisco dimers is polymerization-incompetent and off-the polymerization pathway</u>*.

A mechanism with two distinct Rubisco conformers (R_2 and R_2*), in which the incompetent dimer (R_2^*) is populated through a branched, off-pathway equilibrium can also potentially account for the observed population of Rubisco dimers. The non-productive, polymerization-incompetent dimers can form either through association of two Rubisco monomers (K_{dimer}; Scheme 4a) or isomerization (K_{isom}, Scheme 4b) of a polymerization-competent Rubisco dimer (R_2) according to the following reaction pathways:



Scheme 4a



Scheme 4b

In Scheme 4a and 4b, if the nucleus size $n = 3$, the overall nucleation constant (K_{nuc}) can be defined as $K_{nuc} = [R]^3/[R_3]$ or in a stepwise manner as $K_{nuc} = K_{nuc1}K_{nuc2}$, with $K_{nuc1} = [R]^2/[R_2]$. The non-productive dimer concentration ($[R_2^*]$) is given by $[R_2^*] = [R]^2/K_{dimer}$ in Scheme 4a, and the following relationship holds (Eq. 13a):

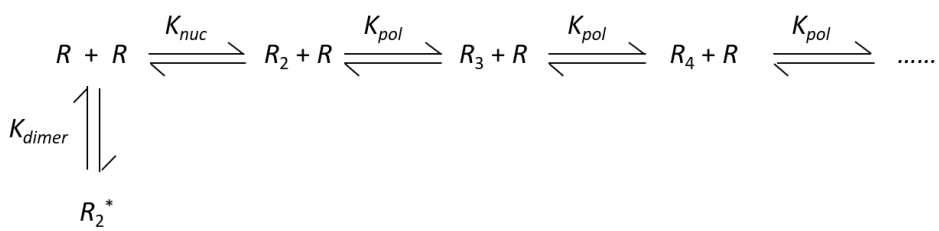
$$\frac{[R_2{}^*] + [R_2]}{[R_3]} = \frac{\left(\frac{1}{K_{nuc1}} + \frac{1}{K_{dimer}}\right)[R]^2}{\frac{[R]^3}{K_{nuc}}} = \frac{(K_{nuc1} + K_{dimer})K_{nuc2}}{[R]K_{dimer}} = \frac{\left(\frac{K_{nuc1}}{K_{dimer}} + 1\right)K_{nuc2}}{[R]} \qquad \text{Eq. 13a}$$

In Scheme 4b, $[R_2{}^*] = [R_2]/K_{isom} = [R]^2/(K_{nuc1}K_{isom})$, and the following relationship holds (Eq. 13b):

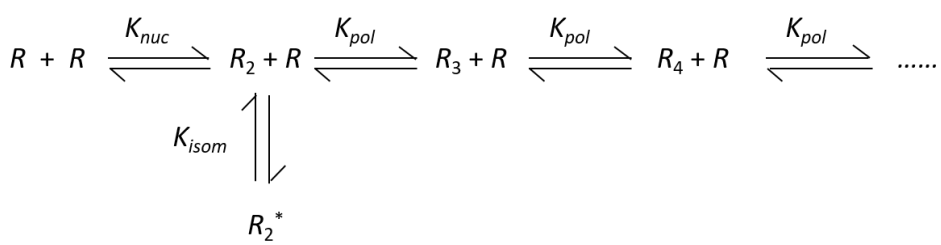
$$\frac{[R_2{}^*] + [R_2]}{[R_3]} = \frac{\left(\frac{1}{K_{nuc1}} + \frac{1}{K_{nuc1}K_{isom}}\right)[R]^2}{\frac{[R]^3}{K_{nuc}}} = \frac{(1 + K_{dimer})K_{nuc2}}{[R]K_{dimer}} = \frac{\left(\frac{1}{K_{dimer}} + 1\right)K_{nuc2}}{[R]} \qquad \text{Eq. 13b}$$

In both non-productive, off-pathway Rubisco dimer pathways, the observed dimer population will exceed the nucleus population (R_3) in the polymer length distribution (Fig. 3) when K_{isom} or K_{dimer} favors R_2* and/or K_{nuc2} is weak and favors R_2 over R_3.

As explained above in the case of an on-pathway, non-productive dimer (Scheme 3b), if the ends of a Rubisco dimer and trimer are identical, K_{nuc2} is defined as K_{pol}, K_{nuc} is defined as K_{nuc1}, and the nucleus size would be a dimer (R_2, n = 2), according to:



Scheme 4c

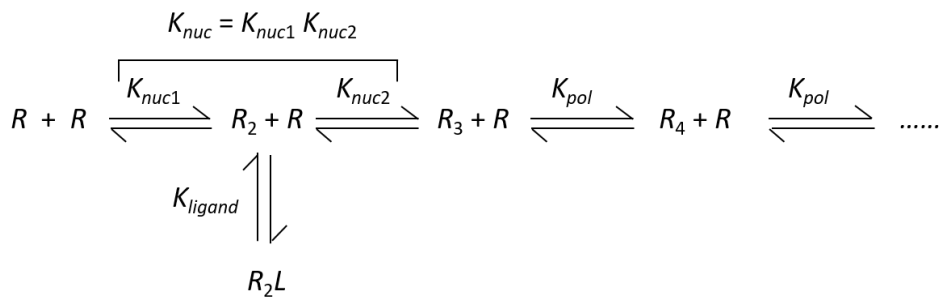


Scheme 4d

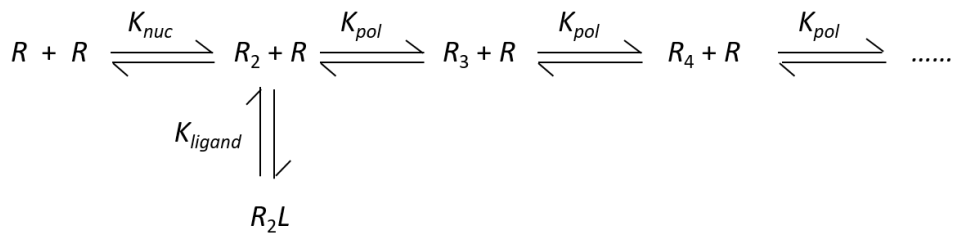
In these two cases, the total observed dimer population will be over counted when K_{nuc} is weaker than K_{dimer}, thereby favoring R_2*, or K_{isom} favors R_2*, as above. As in *Case 2*, such an effect could not alter secondary or tertiary structure, or it would be visible in the structural biology analysis.

Case 4: <u>*A regulatory factor in α-carboxysomes binds and sequesters Rubisco dimers*</u>.

An off-pathway, polymerization-incompetent dimer (R_2^*; *Case 3*) could potentially originate from the binding of a regulatory ligand (*L*), resulting in dimer sequestration. In this case, Schemes 4b and 4d become Scheme 5a and 5b, respectively, and all other relationships hold.



Scheme 5a



Scheme 5b

It is possible that a small-appearing ligand could be invisible in the STA structures if its occupancy was only once per complex, as the symmetry constraint in processing would diffuse its signal. While the STA was confirmed with no symmetry imposed (C1), a small-appearing ligand would have too small a signal to drive alignment.

Free energy changes associated with Rubisco polymerization

The total free energy of a polymer chain ($G_j^{o\prime}$) scales with the total number of subunits (Eq. 10) and thus increases when it elongates. The free energy change of a Rubisco fibril associated with incorporation of one subunit ($\Delta G_{j\to j+1}^{o\prime} = \kappa k_B T$) is < 0.5 $k_B T$ (Eq. 10) for all Rubisco samples evaluated here. A half $k_B T$ of energy change seems small since it is equal to the amount of the thermal energy associated with one degree of freedom, but it is >1200-fold larger than that associated with actin filament elongation ($\Delta G_{j\to j+1}^{o\prime}$ = 0.0004 $k_B T$; (32)). A polymer with larger $\Delta G_{j\to j+1}^{o\prime}$ is thermodynamically and mechanically less stable and more susceptible to spontaneous fragmentation due to thermal and applied forces (45–47).

Therefore, Rubisco polymer chains are considerably thermodynamically and mechanically less stable than actin filaments.

We clarify here for general readers that the two free energy changes for subunit incorporation at polymer ends discussed here, $\Delta G^{0'}{}_{pol}$ vs. $\Delta G^{o'}_{j\to j+1}$, are not equivalent. The term $\Delta G^{o'}_{j\to j+1}$ is the amount of free energy added to a polymer when it extends one subunit, while $\Delta G^{0'}{}_{pol}$ is the free energy change for the overall reaction of a monomer binding to polymer end. Both are material properties, but $\Delta G^{o'}_{j\to j+1}$ is an intrinsic constant, inversely scaling with polymers' thermal and mechanical stabilities, while $\Delta G^{0'}{}_{pol}$ is an intrinsic constant reflecting the feasibility for monomer incorporation at a polymer end.

The (standard) Gibbs free energy change $\Delta G^{0'}{}_{pol}$ is the negative change (loss) given by the difference between the Gibbs free energy ($G^{0'}$) sum of a polymer with j subunits plus a free monomer $\left(G^{0'}{}_{j} + G^{0'}{}_{1}\right)$ and a polymer with j + 1 subunits $\left(G^{0'}{}_{j+1}\right)$:

$$\Delta G^{0'}{}_{pol} = G^{0'}{}_{j+1} - \left(G^{0'}{}_{j} + G^{0'}{}_{1}\right) = k_B T \ln K_{pol} \qquad \text{Eq. 14}$$

It reflects the free energy change associated with subunit incorporation at a filament end and is independent of the length of the elongating filament it is adding to. The polymerized subunit free energy is lower than that of a free monomer in solution, so subunit elongation is associated with a negative Gibbs free energy change.

In contrast, $\Delta G^{o'}_{j\to j+1} = G^{0'}{}_{j+1} - G^{0'}{}_{j}$ is defined as the (standard) free energy increase of the polymer itself when it elongates by one subunit ($\Delta G^{o'}_{j\to j+1} = \kappa k_B$, Eq 10). A longer polymer (of a given subunit type, such that the free energy change per subunit is constant) has a larger total free energy than a shorter one and is less therefore stable.

The two free energy changes refer to different reacting components are related according to the following, obtained by combining Eqs. 10 and 14:

$$\Delta G^{o'}_{j\to j+1} = k_B T\left(\ln\left(K_{pol}\right) - \ln(A_1)\right) = \Delta G^{0'}{}_{pol} - k_B T \ln(A_1) \qquad \text{Eq. 15}$$

or

$$\Delta G^{o'}_{j\to j+1} - \Delta G^{0'}{}_{pol} = -k_B T \ln(A_1) \qquad \text{Eq. 15}'.$$

This relationship means the difference in the polymer free energy ($\Delta G^{o\prime}_{j\rightarrow j+1}$) and the free energy ($\Delta G^{0\prime}{}_{pol}$) for subunit incorporation is determined by the monomer concentration in equilibrium with polymer. The relationship between $\Delta G^{0\prime}{}_{pol}$ and $\Delta G^{o\prime}_{j\rightarrow j+1}$ (Eq. 15 and 15') can be independently obtained directly from the Gibbs free energy calculation (described in the Supplement).

Relation of Rubisco chemical state and polymerization

One known chemical difference between cytoplasm and α-CBs is redox state: the cytoplasm is a reducing environment, while the interior of α-CBs is oxidized (28, 29). We therefore tested whether redox state could affect Rubisco polymerization. Using chemical treatments, we observed that the ordered lattices were often larger and more prevalent in oxidized samples compared to reduced samples, but with only subtle differences in K_{pol} (Figs. 4 and S4; Table 3). This finding emphasizes the importance of quantifying binding rather than using morphology as a guide: the decreased volume of the DTT-treated α-CBs likely affected this observation. α-CBs are purified without chemical reducing agents, and thus we anticipate that the untreated sample should be mildly oxidized as has been previously seen (48). One potential explanation is that the DTT treatment was not sufficient to fully shift the Rubisco polymerization, either due to the inability to revert side-chain oxygen additions or due to the relatively short treatment time tolerated by the α-CB shell.

If the Rubisco complexes exist in two populations within the α-CB (for example, reduced and oxidized), and if both populations polymerize and have different K_{pol} values within the observed A_t range, we would predict two inflection points in the data. Conversely, the sole inflection point observed here indicates that either only one polymerizing species is present, or that two polymerizing species exist but only one has a critical concentration within the observed range of A_t. To test this, we introduced a floating parameter in the fitting program as the fraction of total Rubisco A_t that is either unable to polymerize at all or polymerizes beyond the observed A_t range. Fitting to all the three datasets resulted in this parameter being indistinguishable from zero, indicating that Rubisco polymerizes as a single species and most likely that redox state does not affect intrinsic polymerization capability in these experiments. We note that a case in which two species of Rubiscos exist but copolymerize is also consistent with the data.

For a system in which there is only one uniform polymerization species, α-CBs with the dense morphology (Figure 1B) should have Rubisco concentrations slightly below the K_{pol}. However, a previous study of Rubisco lattice formation found that while the ordered morphology consistently had Rubisco concentrations above what we now know to be the K_{pol}, the dense morphology covered a range of concentrations and could occur at Rubisco concentrations above this threshold(26).

Taken together, these findings suggest that changes to α-CB packing under different redox conditions are likely occurring through the action of accessary proteins, such as CsoS2, rather than through modification of the Rubisco itself (*Case 4* above).

Conclusion

In summary, we have developed a novel tomography-based method for measuring polymerization in situ. In a model test of measuring Rubisco inside α-CBs, we were able to measure Rubisco K_{pol} and K_{nuc} constants inside an incompletely characterized physiological system with multiple protein species present, and at low affinities and high concentrations not typically amenable to in vitro study. We were able to determine that polymerization proceeds through a nucleus, and that redox-based effects on polymerization act through a secondary protein partner. This test case demonstrates the power of combining robust biophysical calculation and modeling with the *in situ* capabilities of cryo-electron tomography.

Theory of Protein Polymerization

Protein polymerization can be described as a highly cooperative, concentration-dependent phase transition, analogous to condensation or crystallization (49). This transition (between monomer and polymer phases) occurs at a protein concentration called the critical concentration for polymerization. When the total protein concentration is less than the critical concentration, the protein exists in the monomeric form (phase 1). At total protein concentrations above the critical concentration, polymer forms (phase 2).

Assembled polymer exists in a reversible equilibrium with monomers, which remain at a concentration approximately equal to the 'critical concentration' such that the total polymer concentration is equal to the protein concentration exceeding the critical concentration. For this reason, the critical concentration is sometimes called the critical monomer concentration. The critical concentration of linear polymers is also approximately equal to the overall equilibrium binding constant for monomer incorporation at filament ends (32, 49). The critical concentration is constant under a defined set of conditions and can be influenced by regulatory proteins, chemical modifications, temperature, and solution conditions, including crowding (8).

The critical concentration is usually determined from the apparent inflection point by linear extrapolation of the linear protein concentration-dependence of polymer and/or monomer concentrations (49, 50). However, the inflection points of our Rubisco binding curves (Fig. 4) are difficult to determine accurately by simple linear extrapolations from the regions before and after the critical concentration because of intrinsic noise inherited from the image processing of tomograms. We, therefore, determine the inflection points and critical concentrations from global fits of the monomer and polymer concentration curves (Fig. 4 and supplementary Fig. S4). We also assess how various polymer assembly models, specifically isodesmic and nucleation (with nucleus n = 2 and 3 Rubisco complexes) with limited length or unlimited length, account for the experimental data.

General polymerization theory

The following reaction scheme is general for the growth (i.e., elongation) of a self-associating linear polymer (40):

$$L_j + L_1 \underset{k_{pol,-}}{\overset{k_{pol,+}}{\rightleftharpoons}} L_{j+1}$$

(Scheme 1),

where L_j is a polymer chain consisting of j (=1, 2, …) subunits and L_1 is a monomer (or "polymer" chain comprised of 1 subunit). This reaction is a reversible with an overall dissociation equilibrium constant (K_{pol}), such that polymer L_j becomes L_{j+1} when a monomer incorporates and L_{j-1} when a single subunit dissociates.

Linear polymer assembly often occurs through the formation of a "nucleus", defined as the oligomer size that binds incoming subunits the same as long polymers (affinity K_{pol}; discussed below). Oligomers smaller than the nucleus size are thermodynamically unstable, are thus populated at low concentrations, and tend to rapidly dissociate. In cases where polymerization proceeds by a nucleation-elongation mechanism, the nucleus is defined as an oligomer of n subunits (n >1):

$$\underbrace{L_1 + L_1 + \ldots + L_1}_{n} \underset{}{\overset{K_{nuc}}{\rightleftharpoons}} L_n \quad n \geq 2$$

(Scheme 2),

where K_{nuc} is the overall dissociation equilibrium constant for nucleus formation. Once a nucleus is formed, polymer elongation follows Scheme 1. In this case, $j \geq n$ in Scheme 1.

Isodesmic polymerization theory

When the nucleus size n = 1, polymerization does not proceed via a nucleation-elongation mechanism; a monomer can directly bind another to form the polymer (Scheme 1). This is a special case called isodesmic polymerization. In this case, Scheme 2 is part of Scheme 1 with $K_{nuc} = K_{pol}$ and can be ignored.

According to established theories of isodesmic polymerization (40, 50), the monomer concentration in equilibrium with stable polymer (chain length $i \geq 2$ subunit) is given by the following relation:

$$A_1 = \frac{K_{pol}A_{tot}}{A_{tot} + \frac{K_{pol}}{2} + \sqrt{\left(A_{tot} + \frac{K_{pol}}{2}\right)^2 - A_{tot}^2}} \rightarrow \begin{cases} \sim A_{tot}, & A_{tot} \ll K_{pol} \\ K_{pol}, & A_{tot} \gg K_{pol} \end{cases} \qquad \text{Eq. 1,}$$

where A_1 = [L_1] is the monomer concentration in reversible equilibrium with polymers, A_{tot} is the total concentration of all the polymerizing macromolecules, and K_{pol} is the polymerization equilibrium constant (Scheme 1). The value of K_{pol} is the critical concentration for polymerization(51).

At equilibrium (Eq. 1), the free monomer concentration (A_1) is approximately equal to the total macromolecule concentration ($A_1 \sim A_{tot}$) when $A_{tot} << K_{pol}$, and levels off when $A_{tot} \sim K_{pol}$, yielding a curved, gradual transition around the value of K_{pol} in a plot of monomer concentration versus total macromolecule concentration (52, 53) (Fig. S4). When $A_{tot} >> K_{pol}$, the free monomer concentration (A_1) infinitely approaches the constant K_{pol} value.

Because the total macromolecule concentration is the sum of monomeric (A_1) and polymerized subunits (A_{pol}; $A_{tot} = A_1 + A_{pol}$), the concentration of polymerized subunits (A_{pol}) is given by:

$$A_{pol} = A_{tot} - A_1 = \sum_{j=2}^{\infty} jA_j = 2A_2 + 3A_3 + \cdots$$

$$= A_{tot}\left(1 - \frac{K_{pol}}{A_{tot} + \frac{K_{pol}}{2} + \sqrt{\left(A_{tot} + \frac{K_{pol}}{2}\right)^2 - A_{tot}^2}}\right) \rightarrow \begin{cases} \sim 0, & A_{tot} \ll K_{pol} \\ A_{tot} - K_{pol}, & A_{tot} \gg K_{pol} \end{cases} \qquad \text{Eq. 2}$$

where A_j = [L_j], j = 1, 2, 3 … is the concentration of polymers with chain length equal to i subunits. The approximated expression (far right of Eq. 2) indicates the polymerized subunit concentration is negligible ($A_{pol} \approx 0$) when the total macromolecule concentration is much less than the polymerization equilibrium constant ($A_{tot} << K_{pol}$), whereas it increases approximately linearly with the total macromolecule concentration ($A_{pol} \approx A_{tot} - K_{pol}$) when it far exceeds the polymerization equilibrium constant ($A_{tot} >> K_{pol}$). This relationship yields K_{pol} as the center of transition in polymer concentration versus total macromolecule concentration plots as well (52, 53). The transitions are gradual rather than abrupt for isodesmic polymerization, because the transitions are not cooperative.

Cooperative nucleation-elongation polymerization theory

Cases of linear polymer assembly in which $n \geq 2$ monomers form a "nucleus" for polymerization are cooperative (i.e., free monomers bind nuclei and polymers longer than nuclei more strongly than other monomers) and are referred to as nucleation-elongation polymerization. In these cases, the equation for monomer concentration (A_1) in reversible equilibrium with polymer is given by (40):

$$\left(1-\frac{A_1}{K_{pol}}\right)^2 (A_t - A_1) - \frac{K_{pol}^n}{K_{nucl}}\left(\frac{A_1}{K_{pol}}\right)^n \left((n-1)\left(1-\frac{A_1}{K_{pol}}\right)+1\right) = 0 \qquad \text{Eq. 3}$$

where the nucleation constant (K_{nuc}) is the overall dissociation equilibrium constant for nucleus formation (Scheme 2). Eq. 3 is an implicit polynomial equation of A_1 and it can be solved numerically for A_1, from which the subunit concentration in all polymers is calculated from mass conservation according to $A_{pol} = A_{tot} - A_1$, yielding:

$$A_{pol} = A_{tot} - A_1 = \sum_{j=n}^{\infty} jA_i = nA_n + (n+1)A_{n+1} + \cdots \qquad \text{Eq. 4.}$$

Eq. 4 treats non-polymer oligomers smaller than the nucleus as "monomers". These intermediates are unstable and exist at low concentrations. Oligomers equal to or longer than the nucleus size are treated as "polymer".

Eqs. 3 and 4 (or Eqs. 1 and 2 for isodesmic polymerization) can used to fit the total macromolecule concentration (A_{tot})-dependence of both monomer (A_1) and polymer populations (A_{pol}). When numerically solving Eq. 3 for A_1, $0 \leq A_1 \leq A_{tot}$ since A_1 is the monomer concentration. On the other hand, a value of $A_1 \geq K_{pol}$, is singularity of Eq. 4 and the summation leading to Eq. 3 (54), so the mathematically allowed solution of polynomial Eq. 3 must be $A_1 < K_{pol}$. Therefore, the combined constraint for A_1 value is in the range $0 \leq A_1 < \min(K_{pol}, A_{tot})$, where the term $\min(K_{pol}, A_{tot})$ means the smaller value of K_{pol} and A_{tot}.

Similar to the isodesmic case, when $A_{tot} << K_{pol}$, i.e., $A_1 < A_{tot} << K_{pol}$, $A_1 \sim A_{tot}$ (Eq. 3), indicating that the free monomer concentration at equilibrium (A_1) increases proportion to the total macromolecule concentration, but approaches the K_{pol} value when $A_{tot} >> K_{pol}$. Also similar to isodesmic polymerization, the polymer subunit concentration (A_{pol}) is negligible when the total macromolecule concentration is low ($A_{tot} << K_{pol}$), and the total polymer concentration (A_{pol}) approaches $A_{pol} - K_{pol}$ when $A_{tot} >> K_{pol}$. Therefore,

K_{pol} is the observed critical concentration (i.e., inflection point) in the plots of monomer/polymer versus total mscromolecule concentration (*Bull. Math. Biol.* **60**, 449–475 (1998)).

Polymerization theory within a confined volume

The polymerization theory above assumes unconstrained polymerization in an infinite accessible volume. The total concentration of macromolecule and polymer length in infinite accessible volume is unlimited. However, the length distribution remains an exponential decay (28,30). When polymerization occurs under confinement, as is the case of Rubisco polymerization inside α-CBs, the maximum concentration of polymerization macromolecules is limited by the exclusion volume of polymerization units. Since the volume excluded by a single Rubisco is ~ $10\times10\times10$ = 1000 nm^3 (Table 2), and the volume of a typical untreated α-CB is ~ 7×10^{-4} μm^3 = 7×10^5 nm^3 (Table 1), the maximum number of Rubisco complexes that can fit into a typical α-CB is given by the ratio of the two volumes (i.e., N_{max} ~ 7×10^5 $nm^3/10^3$ nm^3 = 700, which is equivalent to $700/(6.02\times10^{17}$ $\mu mol^{-1})/7\times10^{-4}$ μm^3 ~ 1700 μM). However, the maximum observed Rubisco concentration (A_{tot}) in our untreated α-CB samples is ~785 μM (Fig. 4A), less than half of the estimated theoretic value of 1700 μM, suggesting that not all the α-CB volume is accessible to Rubisco.

The polymer length in a confined volume is naturally limited by the accessible volume as well. Again in the case of Rubisco polymerization in α-CB, an untreated α-CB with typical volume of 7×10^{-4} μm^3 is equivalent a 55 nm radius ball (see packing section), which is about the size of 5.5 Rubisco complexes lined up (size of a Rubisco ~10 nm, see above), indicating that a maximum of only 10 – 11 polymerized Rubisco complexes can potentially occupy the diameter of a typical untreated α-CB. This is precisely the observed maximal Rubisco chain length (10 – 11 Rubisco complexes) observed in the data for all treated and untreated samples (Table 2). Therefore, confinement appears to contribute to the observed Rubisco polymer distribution. Accordingly, the polymerization theory presented above must be modified to consider confinement and subsequent geometric constraints.

Scheme 1 and 2 above still apply under confinement, but the maximum possible total macromolecule concentration (A_{tot}) and length is limited (L_m). When A_{tot} is limited by confinement before

the polymer length is limited, the theories of isodesimic and nucleation polymerizations are the exactly the same as above (Eq. 1-4), and the analysis remains the same.

In contrast, when confinement limits the polymer length before it limits A_{tot}, a different treatment is needed, which we present here. Defining m as the maximum number of subunits in a polymer chain, then L_m is a polymer with the maximum length and its concentration is $A_m = [L_m]$, which obeys the following equation

$$\frac{dA_m}{dt} = k_{on,pol}A_1A_{m-1} - k_{off,pol}A_m \qquad \text{Eq. 5}$$

where $k_{on,pol}$ and $k_{off,pol}$ are the average on- and off-rate constants for monomers with concentration A_1 to incorporate at the either end. The right hand side of Eq. 5 lacks the terms linking A_{m+1} (see Ref (40)), the concentration for L_{m+1}. Despite this difference, the general solution for the concentration of polymers with j number of subunits is the same as that in an unlimited space (40), i.e. an exponential function of j according to:

$$A_j = \frac{K_{pol}^n}{K_{nuc}}\left(\frac{A_1}{K_{pol}}\right)^j = \frac{K_{pol}^n}{K_{nuc}}e^{-\kappa j}, \qquad j = n, n+1, \dots, m \qquad \text{Eq. 6}$$

Where polymerization constant (or critical concentration) $K_{pol} = \frac{k_{off,pol}}{k_{on,pol}}$ and the exponential decay constant $\kappa = -\ln\left(\frac{A_1}{K_{pol}}\right)$. The total polymerizing macromolecule concentration A_{tot} is equal to the sum of monomer concentration (A_1) and the polymerized subunits (A_{pol}) as follows:

$$A_{tot} = A_1 + \sum_{j=n}^{m} jA_j = A_1 + \frac{K_{pol}^n}{K_{nuc}}\sum_{j=n}^{m} j\left(\frac{A_1}{K_{pol}}\right)^j \qquad \text{Eq. 7}$$

$$= A_1 + \frac{\frac{K_{pol}^n}{K_{nuc}}}{\left(1-\frac{A_1}{K_{pol}}\right)^2}\left(\left(\frac{A_1}{K_{pol}}\right)^n\left(n-(n-1)\left(\frac{A_1}{K_{pol}}\right)\right) - \left(\frac{A_1}{K_{pol}}\right)^{m+1}\left(1+m-m\left(\frac{A_1}{K_{pol}}\right)\right)\right) \qquad \text{Eq. 7}'$$

Eq. 7 is a polynomial equation for the monomer concentration A_1. It has one and only one positive solution in the region $0 < A_1 < A_{tot}$, which can be readily solved numerically for any given m ($> n$), K_{pol} and K_{nuc} values. Therefore, Eq. 7 applies for polymerization under length confinement in a limited space, where Eq. 3 applies above for polymerization in an unlimited space.

Under length confinement, A_1 is obtained by solving Eq. 7, the concentration of polymerized subunits can be calculated using Eq. 4 with the summation up to m instead of infinity, or calculated by $A_{pol} = A_{tot} - A_1$, as presented above for an unlimited space. For isodesmic polymerization in a confined volume with a limited maximum chain length m, Eq. 7 can still be used with $K_{nuc} = K_{pol}$ and the summation starting $n = 2$. Note that for isodesmic polymerization, the integer number of $n = 2$ in Eq. 7 reflects the summation starting number because it doesn't need to form nuclei for polymerization or the number of molecules to form nucleus is $n = 1$ (see above discussion), whereas, for nucleation polymerization ($n \geq 2$), the number n in Eq. 7 means both summation starting and nucleus number.

Eq. 6 is the general expression for concentration of polymers with a given number of subunits (j) for a length-unlimited polymer model as well (40). However, according to Eq. 6, in the monomer concentration region of $A_1 \geq K_{pol}$, the polymer concentration A_j increases with the chain length j subunits, according to Eq. 8:

$$A_{j+1} = \frac{K_{pol}^n}{K_{nuc}}\left(\frac{A_1}{K_{pol}}\right)^{j+1} = \frac{K_{pol}^n}{K_{nuc}}\left(\frac{A_1}{K_{pol}}\right)^{j}\left(\frac{A_1}{K_{pol}}\right) = A_j\left(\frac{A_1}{K_{pol}}\right) > A_j, \qquad j = n, n+1, \ldots \qquad \text{Eq. 8}$$

Longer polymers are thermodynamically less stable than shorter ones because statistically they have potential sites for fragmentation (47). Accordingly, there should be fewer long polymers than short polymers observed in the length distribution. Hence, for thermodynamically stable polymers $A_1 < K_{pol}$ is required in both length-limited and -unlimited models. In the case of a length-unlimited polymer model, substituting A_j from Eq. 6 into Eq. 4 for calculating A_{pol} will not converge if $A_1 \geq K_{pol}$, since A_{pol} approaches infinity (i.e., $A_1 \geq K_{pol}$ is singularity of Eq. 4). Therefore, in a length-unlimited polymer model $A_1 < K_{pol}$ is not only a thermodynamic requirement, but also a mathematical requirement.

For polymerization under polymer length confinement, $A_1 \geq K_{pol}$ is no longer a problem for the polynomial equation Eq. 7. Thus, mathematically, $A_1 \geq K_{pol}$ can be a solution of Eq. 7. In this case, the monomer concentration A_1 is limited only by A_{tot} since it should never be more than A_{tot}, but it can potentially increase with A_{tot} continuously unlike that A_1 has a constant mathematical limit in an unlimited space. Nevertheless, when $A_1 > K_{pol}$, Eq. 6 indicates that the polymer concentration A_j increase with its length exponentially, i.e., the larger the chain length j, the more concentrated the polymer as noted above. This is thermodynamically unfavorable. For this reason, we can still confine $A_1 < K_{pol}$ to conform to the

thermodynamic stability, as we do in this work (see below). When constraining $A_1 \neq K_{pol}$, the polynomial in Eq. 7 can sum up to Eq. 7'. Eq. 7' is another form of Eq. 7 and can be useful in some cases. Since $A_1 = K_{pol}$ is a pseudo singularity point of Eq. 7', Eq. 7' would have problem at the point in programming and care must be taken when using it.

When the maximum chain length m value gets longer (i.e., approaches infinity), the length limited model should behave like a length-unlimited model. This prediction was confirmed using simulated polymerizations with various maximum chain lengths and shared n, K_{pol} and K_{nuc} values (Fig. S3); polymerization with m = 50 (cyan) is very close to the length-unlimited polymer, and essentially identical when m = 100 (magenta).

Both the maximum polymer length (m) and the total macromolecule concentration (A_{tot}) are limited by the accessible volume as discussed above. If A_{tot} reaches a limit first, the maximum polymer chain length can't grow infinitely. Similarly, if the monomer concentration A_1 is limited to K_{pol}, i.e., $A_1 < K_{pol}$ under the thermodynamic stability consideration, all the polymer concentrations are also limited according to Eq. 6, and then, the maximum allowed total macromolecule concentration A_{tot} in the length limited model is also automatically limited (Fig. S3B) by the mass conservation of Eq. 7, i.e.,

$$A_{tot} = K_{pol} + \frac{K_{pol}^n}{K_{nuc}} \sum_{j=n}^{m} j \left(\frac{K_{pol}}{K_{pol}}\right)^j = K_{pol} + \frac{K_{pol}^n}{K_{nuc}} \frac{(m+n)(m-n+1)}{2} \qquad \text{Eq 9,}$$

In the region of $A_1 < K_{pol}$, the simulated polymerizations in Fig. S3B are exactly the same as those in Fig. S3A, but in the region $A_1 > K_{pol}$, the polymerizations in Fig.S3B have no data (Eq. 7 doesn't have $A_1 < K_{pol}$ solution). The longer the maximum allowed chain length m is, the larger the maximum allowed A_{tot} is (Eq. 9 and Fig. S3B). The maximum allowed chain length goes hand to hand with the maximum allowed A_{tot} and when the maximum allowed chain length approaches infinity, so does the maximum allowed A_{tot}.

The mathematical models for polymerization presented here and use in this work (Eqs. 1-9) do not consider polymer fragmentation and annealing, which can potentially affect polymerization and the length distribution (55). However, we assume Rubisco polymer fragmentation and annealing occur much less frequent than other linear polymers (e.g., actin) since they are relatively short by comparison.

Materials and Methods

α-Carboxysome Preparation

We purified α-CBs from wild-type *H. neapolitanus* (56). Briefly, wild-type *H. neapolitanus* cells were grown in DSMZ-68 medium in a 10 L chemostat (pH 6.4, ambient air sparged, 30 °C), harvesting cells every 2–3 days by 6000 × g centrifugation. A ~8 g cell pellet (resulting from 10–15 L of culture) was chemically lysed and purified by centrifugations (12k × g 15 min, 40k × g 30 min) and sucrose gradient (25 mL, 10–50% sucrose step gradient, centrifuged 105k × g 35 min). Quality and purity of sample was analyzed by Coomassie stained SDS-PAGE and negative stain TEM. The final sample had an A280 of roughly 10 AU and was stored at 4 °C in TEMB buffer (10 mM Tris pH 8, 1 mM EDTA, 10 mM MgCl2, 20 mM NaHCO3) until use.

Negative stain imaging of condensate

Rubisco (0.5 μM) of L8S8 and the CsoS2 (2.7 μM) were mixed at room temperature for 1 hour. The sample was diluted 5x and Rubisco and CsoS2 condensates were visualized by negative-stain TEM. Formvar/carbon-coated copper grids were prepared by glow discharge before sample application. The grids were washed with deionized water several times before staining with 2% (w/v) uranyl acetate. Imaging was performed on a JEOL 1200 EX transmission electron microscope.

Sample Conditions and Tomography

10 nm gold colloids (Aurion M-CEM-25512) were mixed with 5% BSA in PBS to passivate the surface, then were buffer-exchanged into TEMB buffer and concentrated ~100x. The α-CB sample was diluted with the BSA-coated gold in a 4:1 ratio. Samples were treated with 5 mM diamide or DTT for one hour, because published work on *Salmonella* BMC shells has established that small compounds penetrate the shell within that time (57). Samples were plunge-frozen onto glow-discharged C-flat 2/2-300 copper TEM grids using a Vitrobot. Imaging was performed at the Pacific Northwest Center for Cryo-Electron Microscopy on a Titan Krios G4 with a Gatan K3 camera and BioQuantum energy filter. Tomograms were collected in SerialEM according to Supplementary Table 2. Frame alignment, 2D image

processing, and tilt alignment were performed in etomo; CTF estimation in ctffind4; and tomogram reconstruction in novaCTF following previously published pipelines (58–61).

Subtomogram averaging

Particles were identified in SIRT-like filtered tomograms using a customized workflow designed to identify every Rubisco complex within each α-CB (26, 30). The estimated false negative and false positive rates were 1–2% of total Rubisco particles within a α-CB when compared with manual identification in a clear, high-defocus tomogram. Unfiltered map resolutions were not degraded by using every single particle identified, further strengthening the evidence for the strong success rates in particle identification. Resolution was calculated using a loose mask in Dynamo; final mask-corrected FSC was calculated in RELION5 using a five-pixel extended mask derived from the low-pass filtered 6UEW PDB structure.

The general subtomogram averaging workflow was based upon a previous high-resolution approach originally derived from the Briggs group (9, 26). Following particle identification, subtomograms were re-extracted from 3D CTF-corrected weighted back projection tomograms. α-CBs were divided into half-sets, and subtomogram alignment and averaging was performed in Dynamo v.1.1.532, using the minimum number of iterations possible at each bin size and employing adaptive bandpass filtering throughout (62). C4 symmetry was tested for accuracy, then imposed after searches were restricted to <90-degree rotations. The best 80% of particles according to cross-correlation were used to generate the reference for the next iteration, but all particles were aligned and participated in the next round. Masks throughout were kept loose. When the resolution ceased to improve by at least 1 Fourier pixel per iteration, the half-maps were sent to RELION5 for post-processing and final resolution calculations (63). Rigid body docking was performed in UCSF Chimera (64).

The reference used for STA was first developed fully *de novo* utilizing a low-resolution average of 124 Rubisco complexes excluded from the dataset. Following a full STA procedure that brought the resolution subnanometer, the *de novo* reference was compared to existing Rubisco structures and confirmed to match the 6UEW structure in all visible helices and bulky side chains. We then performed a second round of STA on the independent half-sets using the adaptive bandpass-filtered 6UEW structure, which refined the resolution further.

The Rubicso-Rubisco interface was generated by isolating all particles in trimers. Averaged particles were shifted within the box to center the interface and a mask was generated to isolate the interface. The model (Fig. 1E,F) showing the stacked Rubisco-Rubisco dimer was generated by fitting the monomer maps into the masked interface volume. The interface (Fig. 1G) is the STA volume generated by Dynamo.

Data Analysis

We used a previously published Matlab script package to determine Rubisco binding states (31). Briefly, the particle position and orientation were used to determine whether two adjacent Rubisco complexes were likely to be bound, using the parameters of binding requiring an inter-particle distance of 9.936nm and angle of 25°. We previously determined that Rubisco complexes immediately adjacent to the shell appear to behave differently (26), and thus allow calculations both including and excluding these particles from analysis (31).

Rubisco Filament Length Distribution Analysis

At equilibrium, the polymer length distribution is an exponential function of A_1 (Eq. 6). On the other hand, the remaining unpolymerized monomer concentration A_1 is function of A_{tot} (Eq. 1, 3, and 7, depending on the polymerization model), and increases with A_{tot}, but $A_1 \rightarrow K_{pol}$ as $A_{tot} \rightarrow \infty$.

Because the polymer length distribution function depends on A_1 (Eq. 6), and A_1 depends on A_{tot} (Eqs. 1,3, and 7), different A_{tot} yield different polymer length distributions. Accordingly, using Eq. 6 to fit the length distribution data to an exponential requires that all data be acquired at the same A_{tot}. Nevertheless, for a nucleation polymerization, polymers appear only when $A_{tot} > K_{pol}$, and in this region, $A_1 \sim K_{pol}$ and is almost constant and changes little with A_{tot} (Fig. 4). Consequently, the polymer length distribution (Eq. 6) changes very little with A_{tot}, and aggregating all the polymer lengths inside different α-CBs with different A_{tot} (Fig 3) is acceptable for a length distribution analysis. Since the exponential decay constant $\kappa = -\ln\left(\frac{A_1}{K_{pol}}\right)$ changes little with A_{tot}, it can be used as a characteristic property of polymerization, which can be compared among different macromolecules, as done previously(32).

Polymerization Analysis

The total Rubisco concentration-dependence of the monomer- and polymer- (including dimer for n = 2) subunit concentrations were globally fitted to the chain length limited and unlimited isodesmic (Eqs. 1 and 2 for length unlimited; Eq. 7 with $K_{nuc} = K_{pol}$ for length limited) and n =2 cooperative nucleation-elongation (Eqs. 3 and 4 for length unlimited; Eq. 7 for length limited) polymerization models, whereas the total Rubisco concentration-dependence of the monomer+dimer- and polymer- (excluding dimer) were globally fitted to the chain length limited and unlimited n =3 cooperative nucleation-elongation (Eqs. 3 and 4 for length unlimited; Eq. 7 for length limited) polymerization models. The fitting was performed using Origin software (OriginLab, MA, USA; www.OriginLab.com). The fitting procedure was controlled by a user-defined parameter (poly_switch) that fits each data set, unbound or bound, to a different equation, with some parameters shared among the equations. When poly_switch = 0, the unbound species, monomer concentration (A_1) or monomer+dimer concentration is fitted by solving Eq 1 (isodesmic), or Eq. 3 or Eq. 7 (nucleation); when poly_switch = 1, the polymer subunit concentration (A_{pol}) is fitted using $A_{tot} - A_1$ with the monomer value (A_1) obtained from Eq. 1 or Eq. 3 or Eq. 7, respectively. Another parameter, non_poly, was introduced in the Origin fitting program as the faction of the species that is not able to polymerize in the observed A_{tot} range. In fitting, the combined monomer concentration is calculated as non_poly*A_{tot} +A_1 and the A_1 of polymerization species is determined with the polymerization species (1 – non_poly)* A_{tot}. The analyses of all three treated and untreated Rubisco polymerization show non_poly = 0 within the uncertainties.

The simulated data according to the four polymerization models were done with Matlab (MathWorks, MA, USA; www.mathworks.com)

Eq. 3 and Eq. 7 are implicit equations of the monomer concentration A_1. It was solved for A_1 in fitting procedures and simulations with a custom numerical routine based on a root-finding algorithm implementing the Newton-Raphson method combined with the Bisect method (65).

In the polymer length limited analyses, the length m = 10 for untreated and DTT treated, but 11 for diamide treated, we also confined the monomer concentration $A_1 \leq K_{pol}$, but let K_{pol} and K_{nuc} values unconstrained.

Acknowledgements

We thank Keith Meyers, Morgan Stephens, and Izabella Soswa for their assistance aligning tomograms and processing data for subtomogram averaging. Molecular graphics and analyses performed with UCSF Chimera (64).

Authors' contributions

L.A.M. and E.M.D.L.C. conceived and supervised the project. K.R., W.C., L.A.M., and E.M.D.L.C. wrote the manuscript. W.C. and K.R. prepared figures. All authors contributed to the analysis and interpretation of the data. L.O. and D.F.S. provided the purified carboxysome samples and contributed images of the Rubisco/csoS2 condensate. L.A.M. prepared the samples for imaging and collected the tilt stacks. K.R., R.H.G., and L.A.M. processed the tomograms, conducted the STA, and carried out the biophysical processing of the particles. W.C. and E.M.D.L.C. analyzed the biophysical results, established the mathematical theory and models, and interpreted the binding results. K.R. finalized the structures and the EMDB/EMPIAR deposition.

Funding

This work was supported by funding from the Ralph W. and Grace M. Showalter Research Trust and NSF CAREER award 2439291 to L.A.M., and funding from NIH NIGMS award R35 GM136656 to E.M.D.L.C. K.R. was funded in part by NIH NIGMS F32 Postdoctoral Fellowship (F32 GM161096). This publication was made possible by an NIGMS-funded predoctoral fellowship to R.H G. (T32 GM132024). A portion of this research was supported by NIH grant R24GM154185 and performed at the Pacific Northwest Center for Cryo-EM (PNCC) with assistance from Vamseedhar Rayaprolu. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS, NIH, or NSF.

Supplementary Materials:

Morphological analysis

Table S1A: Manual morphology classification of diamide-treated α-CBs

Analyst	# Sparse	# Dense	# Ordered
1	35	55	234
2	42	111	176
3	32	173	111
Percentage	11 +- 2 %	35 +- 18 %	54 +- 19 %

Table S1B: Manual morphology classification of DTT-treated α-CBs

Analyst	# Sparse	# Dense	# Ordered
1	5	6	256
2	5	126	137
3	9	144	109
Percentage	2 +- 1 %	34 +- 28 %	63 +- 29 %

Each analyst was instructed to classify morphology as sparse (>33% empty interior volume), ordered (>3 fibrils with >3 Rubisco subunits in one bundle), or dense (not ordered or sparse). Analysts needed to search in three-dimensional space for Rubisco arrangements, and personal biases are strong: for example, in DTT and diamide treated α-CBs (Supplementary Table 1A and B), analyst #1 identifies patterns readily and classified most α-CBs as ordered, while analyst #3 struggled to classify α-CBs without a strong central lattice and thus always found more dense than ordered morphologies. However, analysts were not told what to expect from the datasets, nor were they allowed to see each other's results. The morphology classification Table for untreated α-CBs was reported previously (26).

Many of the discrepancies in the morphology classification arise when α-CBs contain multiple short chains, which can be difficult to identify manually (Supplemental Fig. S1)

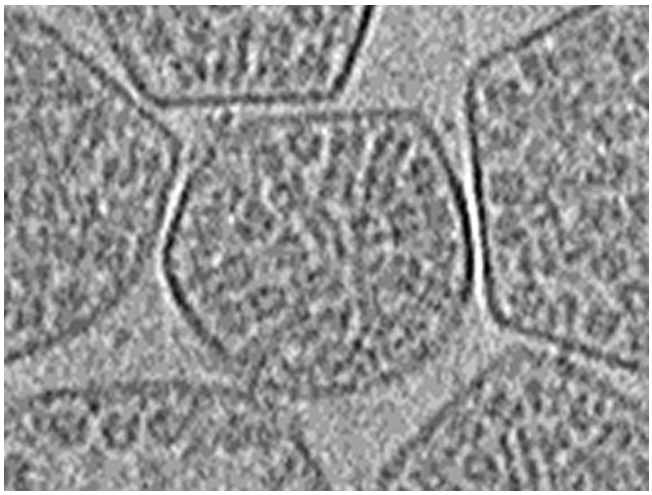


Figure S1. Fibril arrangement in representative α-CB. Representative ambiguous DTT-treated α-CB that contains multiple short fibrils arranged radially.

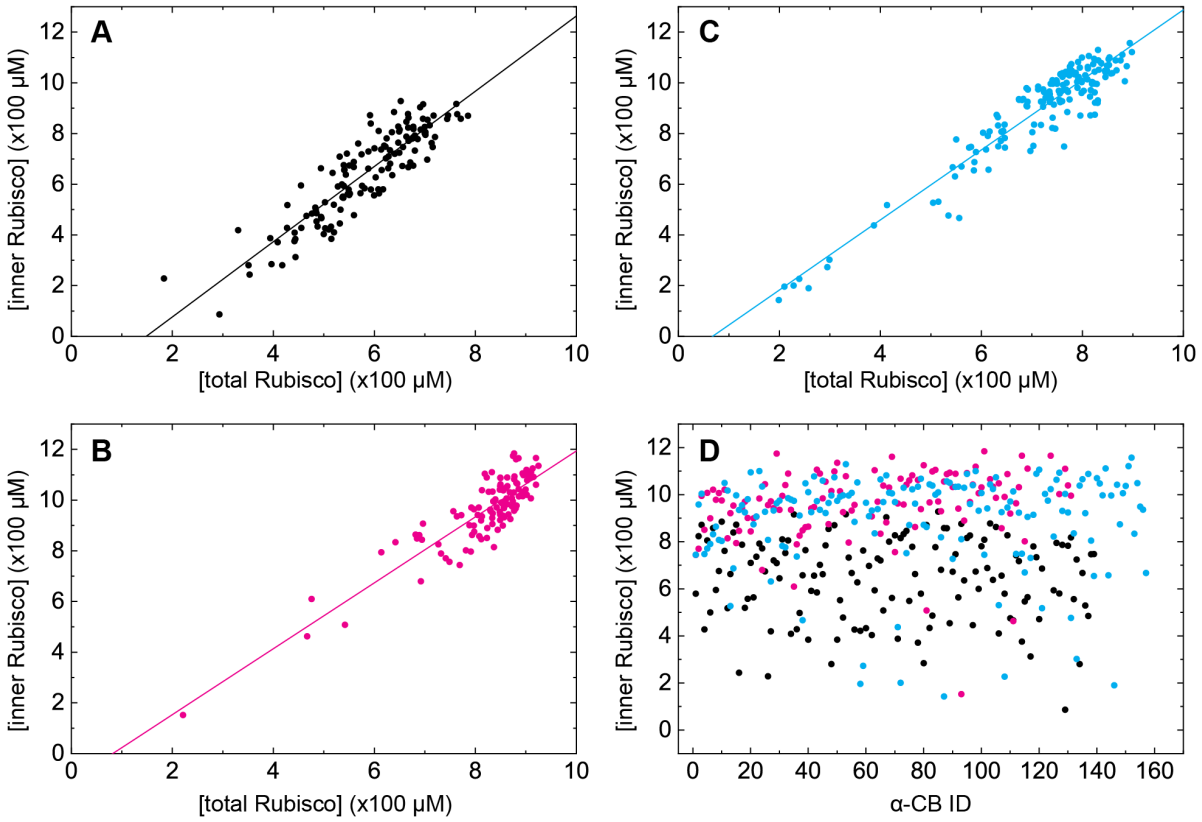


Figure S2. Inner Rubisco concentration. A-C. Correlations of inner Rubisco concentration vs. total Rubisco concentration for untreated, DTT-treated, and diamide-treated in α-CBs. All three panels (**A-C**) are plotted in the same *x*- and *y*-axis scales for easy visual comparison of the slopes of line fits (straight lines through the data). **D.** Comparison of inner Rubisco concentrations for untreated (black), and DTT treated (magenta) and diamide treated (blue) samples. Data in all panels are summarized in Table 1.

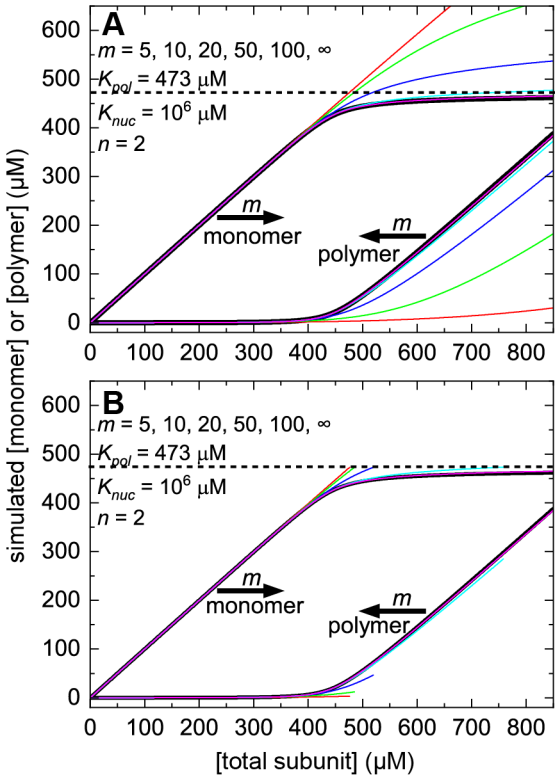


Figure S3. Simulation of maximum chain length (*m*) dependence of nucleation polymerization curves. A. Total subunit concentration A_{tot} dependence of monomer A_1 and polymer A_{pol} Rubisco (dimer and longer) concentrations calculated by solving the maximum chain length limited model (Eq. 7) for m = 5 (red), 10 (green), 20 (blue), 50 (cyan) and 100 (magenta) and chain length unlimited model (Eq. 3, equivalent to $m = \infty$, thick black). The other parameter values: K_{pol}, K_{nuc}, and $n = 2$ are same for all and indicated on the plot. The horizontal arrows with m above them indicate the curve shape trend with increasing m value. The horizontal thin dashed line indicates K_{pol} where the monomer concentration exceeding it, $A_1 > K_{pol}$, is thermally unfavorable. For the length unlimited nucleation model (black line), A_1 is always less than, but approaches K_{pol}. On the other hand, the length limited nucleation model can have $A_1 > K_{pol}$ when A_{tot} is large enough. The curve trend shows when $m = 100$ (magenta line), the polymerization curves of limited length model are almost the same as those of length unlimited model (black line). **B.** The simulated curves are the same curves as those in panel **A** except confining $A_1 \leq K_{pol}$, such that the region $A_1 > K_{pol}$ of all simulated monomer curves in Panel A have been cut off by the horizontal thin dashed line. When confining $A_1 \leq K_{pol}$, the subunit concentration in polymer form also confined, therefore, the polymer curves are also cut off. The purpose of Panel B is to show which part of a simulated curve for a limited length polymerization model has the monomer concentration $A_1 > K_{pol}$ by comparing the two panels.

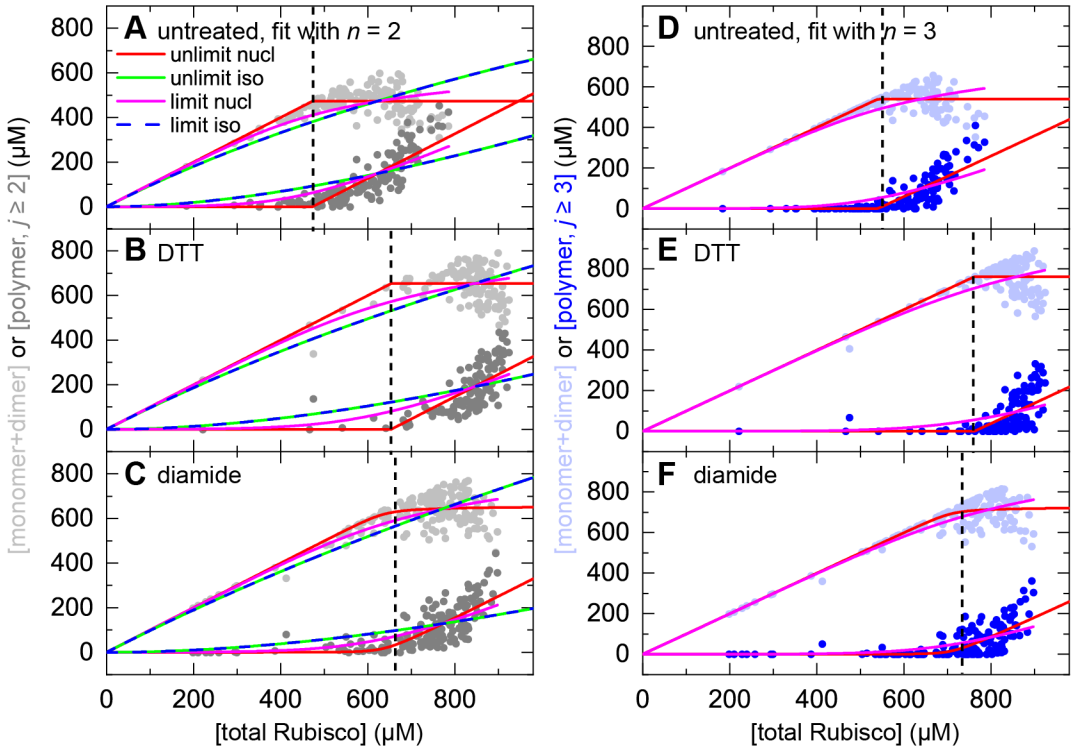


Figure S4. Comparison different fitting model. A-C. Plots of the total Rubisco concentration dependence of unbound Rubisco concentration composed of monomers only (light gray filled circles) and bound Rubisco concentration in dimer and longer polymers (gray filled circles). The unbound and bound data are globally fitted with a length unlimited nucleation (n = 2, red lines), length unlimited isodesmic (green lines), length limited nucleation (n = 2, magenta lines) or length limited isodesmic (blue dashed lines) model in untreated, DTT-treated, and diamide-treated in α-CBs. The best global fit result by length limited isodesmic (blue dashed lines) model is basically the same as that by length unlimited isodesmic (green lines), and the two fit lines are on top one another for all three treated and untreated sample (A-C, Table 3). **D-F.** Total Rubisco concentration dependence of unbound Rubisco concentration in monomers and dimers (light blue filled circles) and bound Rubisco concentration in length ≥ 3 polymers (blue filled circles). The data are globally fitted with a length unlimited nucleation (n = 3, red lines) or length limited nucleation (n = 3, magenta lines) model in untreated, DTT- treated, and diamide-treated in α-CBs. In all fits, the unbound concentration A_1 (monomer only in n = 2 nucleation and isodesmic models, whereas monomer+dimer n = 3 nucleation model) is constrained as $A_1 < K_{pol}$. In all length limit models, the maximum chain length is 10 Rubisco complexes for untreated and DTT treated, and 11 Rubisco complexes for diamide treated. The magenta fit lines for nucleation with length limit in all panels stop and don't go all the way to the right as other fitting models do. It is because in the A_t region beyond stopping point, there is no $A_1 < K_{pol}$ solution in Eq. 7. All the fitting results are listed in Table 3 for comparison.

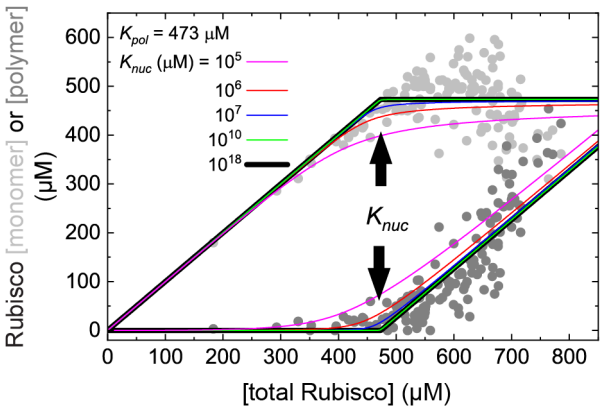


Figure S5. Polymerization inflection point. A large K_{nuc} value increases the sharpness of transition at the inflection point. The monomer (light filled gray circles) and polymer (including dimers, gray filled circles) Rubisco complexes are the same data as untreated Rubisco complexes in Fig. S4A. The smooth thick black lines are the global fits of the length unlimited model to the data with n = 2 (Table 3). The other smooth lines are simulations with the same n and K_{pol}, but K_{nuc} = 10^5 (magenta), 10^6 (red), 10^7 (blue) and 10^{10} (green) □M. The up and down arrows on the plot indicate the direction of simulated monomer (up) and polymer (down) lines approaching with increasing K_{nuc} values at the transition. The larger the K_{nuc} value is, the sharper the transition is. When K_{nuc} = 10^{10} (green), the simulated lines are not much different from the global fitting lines (thick black) to the data.

Alternative Free Energy Proof

In Eq. 15 or 15' of the main manuscript, we obtain the relation of the two standard free energy changes from three different approaches, i.e., one from the solution (Eq. 6) of the kinetic equations associated with the general polymerization schemes Scheme 1 and 2, another one from the standard free energy change for binding ($\Delta G^{0'}_{pol}$, Eq. 11 and 14) and the third from a separated thermodynamic study of polymerization for polymer free energy increase for elongating one subunit ($\Delta G^{o\prime}_{j\rightarrow j+1}$, Eq. 10; (33)). This relationship between $\Delta G^{0'}_{pol}$ and $\Delta G^{o\prime}_{j\rightarrow j+1}$ (Eq. 15 and 15') can be independently obtained directly from the Gibbs free energy calculation. The free energy of a population of *j*-subunits-long polymer (G_j) scales with its concentration (A_j) according to (66):

$$G_j = G^{0'}_j + k_B T \ln\left(A_j\right) \qquad 16.$$

At equilibrium, the free energy change ($\Delta G_{j\rightarrow j+1}$) associated with subunit incorporation (i.e. polymer of *j* subunits elongating to *j* + 1 subunits) is 0 (66), as given by:

$$\Delta G_{j\rightarrow j+1} = G_{j+1} - G_j = G^{0'}_{j+1} - G^{0'}_j + k_B T \ln\left(A_{j+1}\right) - k_B T \ln\left(A_j\right) = 0 \qquad 17$$

and can be rearranged to:

$$\Delta G^{o'}_{j\rightarrow j+1} = G^{0'}_{j+1} - G^{0'}_j = k_B T \ln\left(\frac{A_j}{A_{j+1}}\right) = k_B T \ln\left(\frac{A_j A_1}{A_{j+1}}\frac{1}{A_1}\right) = k_B T \ln\left(\frac{K_{pol}}{A_1}\right) = \Delta G^{0'}_{pol} - k_B T \ln(A_1) \qquad 18$$

Therefore, Eq. 18 is Eq. 15, and it can be independently obtained directly from the Gibbs free energy as defined in college textbooks.

References

1. K. A. Johnson, R. S. Goody, The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).

2. R. P. Bhattacharyya, A. Reményi, B. J. Yeh, W. A. Lim, Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).

3. G. D. Stormo, D. S. Fields, Specificity, free energy and information content in protein–DNA interactions. *Trends Biochem. Sci.* **23**, 109–113 (1998).

4. J. A. Marsh, S. A. Teichmann, Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 551–575 (2015).

5. T. D. Pollard, A Guide to Simple and Informative Binding Assays. *Mol. Biol. Cell* **21**, 4061–4067 (2010).

6. I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, How to measure and evaluate binding affinities. *eLife* **9**, e57264 (2020).

7. A. Hromić-Jahjefendić, H. Krnjić, V. N. Uversky, "Effects of Molecular Crowding on the Structural, Dynamic, and Functional Properties of Biological Macromolecules: A General Overview" in *(Macro)Molecular Crowding: Life of the Pottage*, V. N. Uversky, Ed. (Springer Nature Switzerland, 2025), pp. 15–39.

8. K. B. Frederick, D. Sept, E. M. De La Cruz, Effects of solution crowding on actin polymerization reveal the energetic basis for nucleotide-dependent filament stability. *J. Mol. Biol.* **378**, 540–550 (2008).

9. F. K. M. Schur, *et al.*, An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506–508 (2016).

10. J. Hutchings, V. Stancheva, E. A. Miller, G. Zanetti, Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nat. Commun.* **9**, 4154 (2018).

11. T. A. M. Bharat, *et al.*, Structure of the hexagonal surface layer on Caulobacter crescentus cells. *Nat. Microbiol.* **2**, 17059 (2017).

12. J. A. G. Briggs, Structural biology in situ--the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**, 261–267 (2013).

13. W. Wan, J. A. G. Briggs, "Cryo-Electron Tomography and Subtomogram Averaging" in *Methods in Enzymology*, (Elsevier, 2016), pp. 329–367.

14. G. S. Espie, M. S. Kimber, Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynth. Res.* **109**, 7–20 (2011).

15. B. D. Rae, B. M. Long, M. R. Badger, G. D. Price, Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).

16. J. M. Shively, F. Ball, D. H. Brown, R. E. Saunders, Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacillus neapolitanus. *Science* **182**, 584–586 (1973).

17. G. C. Cannon, *et al.*, Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361 (2001).

18. J. M. Shively, G. L. Decker, J. W. Greenawalt, Comparative ultrastructure of the thiobacilli. *J. Bacteriol.* **101**, 618–627 (1970).

19. Z. Dou, *et al.*, CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J. Biol. Chem.* **283**, 10377–10384 (2008).

20. L. Reinhold, R. Kosloff, A. Kaplan, A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Can. J. Bot.* **69**, 984–988 (1991).

21. L. M. Oltrogge, *et al.*, Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).

22. F. Cai, *et al.*, Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life Basel Switz.* **5**, 1141–1171 (2015).

23. G. C. Cannon, *et al.*, Organization of Carboxysome Genes in the Thiobacilli. *Curr. Microbiol.* **46**, 0115–0119 (2003).

24. N. M. Mangan, A. Flamholz, R. D. Hood, R. Milo, D. F. Savage, pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci.* **113**, E5354–E5362 (2016).

25. H. Wang, *et al.*, Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).

26. L. A. Metskas, *et al.*, Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* **13**, 4863 (2022).

27. T. Ni, *et al.*, Structure and assembly of cargo Rubisco in two native α-carboxysomes. *Nat. Commun.* **13**, 4299 (2022).

28. N. C. Hill, J. W. Tay, S. Altus, D. M. Bortz, J. C. Cameron, Life cycle of a cyanobacterial carboxysome. *Sci. Adv.* **6**, eaba1269 (2020).

29. C. A. Huffine, C. Fontana, A. Avramov, C. Sempeck, J. C. Cameron, Cyanobacteria form a procarboxysome-like structure in response to high CO2. [Preprint] (2024). Available at: https://www.biorxiv.org/content/10.1101/2024.06.28.601118v1 [Accessed 26 November 2024].

30. L. A. Metskas, R. Wilfong, G. J. Jensen, Subtomogram averaging for biophysical analysis and supramolecular context. *J. Struct. Biol. X* **6**, 100076 (2022).

31. F. Lee, *et al.*, A toolkit to characterize protein polymerization from cryo-electron tomography data. [Preprint] (2025). Available at: https://www.biorxiv.org/content/10.1101/2025.04.22.650024v1 [Accessed 3 December 2025].

32. W. Cao, T. E. Sladewski, A. T. Heaslip, E. M. De La Cruz, Bending stiffness of Toxoplasma gondii actin filaments. *J. Biol. Chem.* **301**, 108101 (2025).

33. A. S. Jermyn, W. Cao, W. A. Elam, E. M. De La Cruz, M. M. Lin, Directional allosteric regulation of protein filament length. *Phys. Rev. E* **101**, 032409 (2020).

34. L. Edelstein-Keshet, G. B. Ermentrout, Models for the length distributions of actin filaments: I. Simple polymerization and fragmentation. *Bull. Math. Biol.* **60**, 449–475 (1998).

35. F. Oosawa, *Thermodynamics of the polymerization of protein* (London ; New York : Academic Press, 1975).

36. "Cryo-Electron Tomography and Subtomogram Averaging" in *Methods in Enzymology*, (Academic Press, 2016), pp. 329–367.

37. W. Zhou, *et al.*, Cryo-EM structure-based selection of computed ligand poses enables design of MTA-synergic PRMT5 inhibitors of better potency. *Commun. Biol.* **5**, 1–15 (2022).

38. J. H. Van Drie, L. Tong, Cryo-EM as a powerful tool for drug discovery. *Bioorg. Med. Chem. Lett.* **30**, 127524 (2020).

39. Y. Sun, *et al.*, Decoding the Absolute Stoichiometric Composition and Structural Plasticity of α-Carboxysomes. *mBio* **13**, e0362921 (2022).

40. L. Edelstein-Keshet, G. B. Ermentrout, Models for the length distributions of actin filaments: I. Simple polymerization and fragmentation. *Bull. Math. Biol.* **60**, 449–475 (1998).

41. W. Dai, *et al.*, Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. *J. Mol. Biol.* **430**, 4156–4167 (2018).

42. T. Ni, *et al.*, Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. *Nat. Commun.* **14**, 5512 (2023).

43. L. M. Oltrogge, A. W. Chen, T. Chaijarasphong, J. B. Turnšek, D. F. Savage, α-Carboxysome Size Is Controlled by the Disordered Scaffold Protein CsoS2. *Biochemistry* **63**, 219–229 (2023).

44. R.-Q. Zhou, *et al.*, Structure and assembly of the α-carboxysome in the marine cyanobacterium Prochlorococcus. *Nat. Plants* 1–12 (2024). https://doi.org/10.1038/s41477-024-01660-9.

45. A. C. Schramm, *et al.*, Actin Filament Strain Promotes Severing and Cofilin Dissociation. *Biophys. J.* **112**, 2624–2633 (2017).

46. A. C. Schramm, G. M. Hocky, G. A. Voth, J.-L. Martiel, E. M. D. L. Cruz, Plastic Deformation and Fragmentation of Strained Actin Filaments. *Biophys. J.* **117**, 453–463 (2019).

47. J. P. Bibeau, *et al.*, Twist response of actin filaments. *Proc. Natl. Acad. Sci.* **120**, e2208536120 (2023).

48. W. B. Carpenter, *et al.*, Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution. *J. Phys. Chem. Lett.* **13**, 4455–4462 (2022).

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

50. J. Howard, *Mechanics of motor proteins and the cytoskeleton* (Sinauer Associates, 2001).

51. L. Edelstein-Keshet, G. B. Ermentrout, Models for the length distributions of actin filaments: I. Simple polymerization and fragmentation. *Bull. Math. Biol.* **60**, 449–475 (1998).

52. M. M. J. Smulders, *et al.*, How to distinguish isodesmic from cooperative supramolecular polymerisation. *Chem. Weinh. Bergstr. Ger.* **16**, 362–367 (2010).

53. K. M. Skillman, *et al.*, The unusual dynamics of parasite actin result from isodesmic polymerization. *Nat. Commun.* **4**, 2285 (2013).

54. L. Edelstein-Keshet, G. B. Ermentrout, Models for the length distributions of actin filaments: I. Simple polymerization and fragmentation. *Bull. Math. Biol.* **60**, 449–475 (1998).

55. E. Andrianantoandro, L. Blanchoin, D. Sept, J. A. McCammon, T. D. Pollard, Kinetic mechanism of end-to-end annealing of actin filaments1. *J. Mol. Biol.* **312**, 721–730 (2001).

56. A. I. Flamholz, *et al.*, Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. *eLife* **9**, e59882 (2020).

57. D. S. Trettel, W. Resager, B. M. Ueberheide, C. C. Jenkins, W. C. Winkler, Chemical probing provides insight into the native assembly state of a bacterial microcompartment. *Structure* **30**, 537-550.e5 (2022).

58. B. Turoňová, F. K. M. Schur, W. Wan, J. A. G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J. Struct. Biol.* **199**, 187–195 (2017).

59. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

60. D. N. Mastronarde, Dual-axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 (1997).

61. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).

62. D. Castaño-Díez, M. Kudryashev, M. Arheit, H. Stahlberg, Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012).

63. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

64. T. D. Goddard, C. C. Huang, T. E. Ferrin, Visualizing density maps with UCSF Chimera. *J Struct Biol* **157**, 281–7 (2007).

65. W. H. Press, Ed., *Numerical recipes: the art of scientific computing*, 3rd ed (Cambridge University Press, 2007).

66. D. Voet, J. G. Voet, *Biochemistry*, 4th Ed. (New York : J. Wiley & Sons, 2004).