Impact of Linker Length on Biomolecular Condensate Formation

Trevor GrandPre,[1,2,3] Yaojun Zhang,[2,4] Andrew G. T. Pyo,[1,2] Benjamin Weiner,[1,2] Je-Luen Li,[5] Martin C. Jonikas,[6,7] and Ned S. Wingreen[2,3,7,8,*]

[1]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
[2]*Center for the Physics of Biological Function, Princeton University, Princeton, New Jersey 08544, USA*
[3]*Princeton Center for Theoretical Science, Princeton University, Princeton, New Jersey 08544, USA*
[4]*Department of Physics and Astronomy and Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA*
[5]*D. E. Shaw Research, LLC, New York, New York 10036, USA*
[6]*Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA*
[7]*Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA*
[8]*Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA*

(Received 2 July 2023; accepted 28 November 2023; published 26 December 2023)

Biomolecular condensates are membraneless organelles formed via phase separation of macromolecules, typically consisting of bond-forming "stickers" connected by flexible "linkers." Linkers have diverse roles, such as occupying space and facilitating interactions. To understand how linker length relative to other lengths affects condensation, we focus on the pyrenoid, which enhances photosynthesis in green algae. Specifically, we apply coarse-grained simulations and analytical theory to the pyrenoid proteins of *Chlamydomonas reinhardtii*: the rigid holoenzyme Rubisco and its flexible partner EPYC1. Remarkably, halving EPYC1 linker lengths decreases critical concentrations by tenfold. We attribute this difference to the molecular "fit," i.e., the number of stickers of EPYC1 that can bind to a single Rubisco given the constraint of EPYC1 linker length. We find an inverse relationship between molecular fit and the tendency of EPYC1 and Rubisco to phase separate. Moreover, by computationally varying Rubisco sticker locations we discover that the naturally occurring sticker locations yield the poorest fit for all EPYC1 linker lengths; thus the natural locations optimize phase separation. Surprisingly, shorter linkers mediate a transition to a gas of rods as Rubisco stickers approach the poles. These findings illustrate how intrinsically disordered proteins affect phase separation through the interplay of molecular length scales.

DOI: 10.1103/PRXLife.1.023013

I. INTRODUCTION

Biomolecular condensates—organelles without membranes—are used by cells to organize and orchestrate a multiplicity of processes, ranging from signaling [1–3] and gene regulation [4,5] to metabolism [6] and photosynthesis [7]. The physical properties of these condensates and thus their functions are sensitive to the microscopic features of their constituent biomolecules. Importantly, these microscopic parameters are subject both to evolutionary adaptation via mutation of molecular sequences and to active regulation through chemical modifications. While the interactions that lead to biomolecular phase separation are typically complex, the phase-separated algal organelle called the pyrenoid, which is the locus of carbon fixation, presents a tractable model system in which the interactions are well understood [8–12]. In the leading model alga, *Chlamydomonas reinhardtii*, the multivalency of the two dominant components of the pyrenoid, the rigid Rubisco holoenzyme and its disordered protein EPYC1, are sufficient to drive phase separation *in vitro*, with each Rubisco binding multiple EPYC1s and vice versa [9]. In other Rubisco condensates such as carboxysomes found in alpha and beta cyanobacteria [7,13,14] other multivalent linker proteins mediate phase separation and the associated stickers on Rubisco are at different locations than in *C. reinhardtii*. Here, we focus on the question of how the microscopic "fit," i.e., the number of stickers of EPYC1 that can readily bind to a single Rubisco molecule, influences phase separation.

The interacting domains of phase separating biomolecules (aka "stickers") are typically connected by flexible domains which we refer to in this study as "linkers" (aka "spacers"). For our purposes, linkers are flexible regions that do not engage in strong bonding interactions. Thus linkers are distinct from intrinsically disordered regions (IDRs), within which there can be both stickers and linkers. While much attention has been focused on which domains or residues constitute stickers, the linkers also play key roles in phase separation. For example, in experiments the lengths of linkers and their charge content can substantially affect phase separation [15–17], and influence material coefficients such as viscosity [18,19] and viscoelasticity [20]. Linkers also play functional roles, e.g., by recruiting clients such as kinases [21] and other cargo molecules [22].

*wingreen@princeton.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

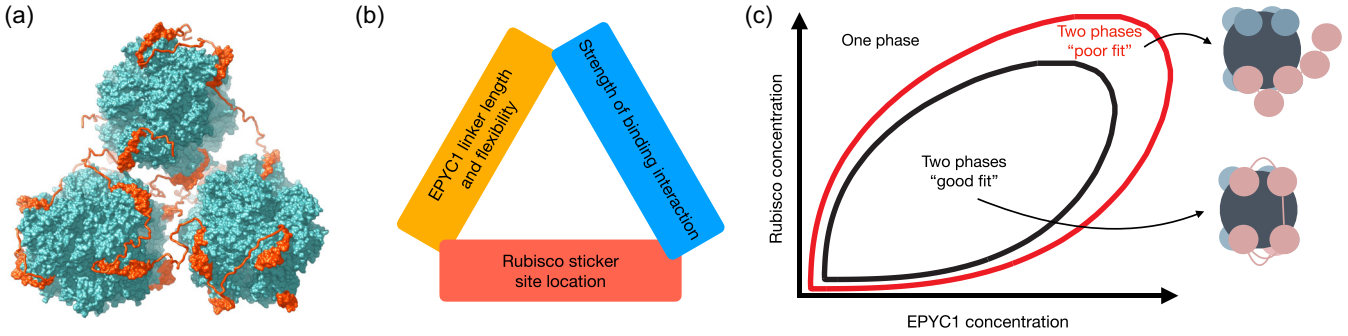

FIG. 1. Dependence of Rubisco-EPYC1 phase separation on microscopic aspects. (a) Illustration of three Rubisco holoenzymes held together by several EPYC1s. The Rubisco holoenzymes and EPYC1 sticker regions were obtained from cryoelectron tomography and the EPYC1 linkers were interpolated based on the correct contour length [9,10]. (b) Three tunable parameters that underpin Rubisco-EPYC1 condensate properties are EPYC1 linker length and flexibility, Rubisco sticker location, and the strength of the binding interaction between EPYC1 and Rubisco stickers. (c) Illustration of how microscopic properties of EPYC1 (i.e., linker length) and of Rubisco (i.e., sticker location) in (b) can control the phase diagram, specifically by giving rise to a good fit, that is, EPYC1 with long linkers can bind almost all its stickers to a single Rubisco molecule, or a poor fit, that is, EPYC1 with short linkers is not able to bind all its stickers to a single Rubisco. In the good fit and poor fit illustrations, the EPYC1 linker length is different, but the Rubisco molecules are identical (see Fig. 2 for more details).

The influence of linker properties on phase separation is confirmed by theory [23–27]. In particular, models with explicit linkers suggest that the volume occupied by linkers within condensates can inhibit phase separation [28], with a linker's effective volume depending on its solubility [23,24]. The flexibility of linkers can also affect phase separation, as seen in models of linker-colloid systems [29–33]. The role of linkers can be modeled implicitly via an effective springlike interaction between adjacent sticker domains [34–36]. For example, an implicit linker model of RNA has shown that phase separation depends on the effective length of linkers due to the loop-entropy cost of forming sticker bonds [37].

A remaining open question is how the length of linkers compared to other relevant length scales may affect phase separation. To address this question, we focus on the Rubisco-EPYC1 system, in which the separation between stickers on the rigid Rubiscos is comparable to the effective linker length between EPYC1 stickers. First, we develop a coarse-grained molecular-dynamics model for Rubisco-EPYC1 phase separation based on experimental measurements of protein sizes, sticker number and location, and dissociation constant. In our work, we keep the number of stickers of Rubisco and EPYC1 fixed, and vary EPYC1 linker length and Rubisco sticker location. We find that EPYC1 linker length drastically changes the phase diagram, with shorter linkers favoring the condensed phase. Next, guided by an analytical dimer-gel theory, we trace the sensitivity to linker length to the fit between a single EPYC1 and a single Rubisco. We confirm this conclusion by computationally varying the location of Rubisco stickers, finding the actual locations to be nearly optimal for phase separation. Moving the stickers toward the poles can even lead to an alternative phase in which Rubiscos form one-dimensional rods. Our results suggest that the sensitivity of phase separation to the geometry of binding may more generally provide strong selection pressure on the lengths of IDRs.

II. RESULTS

Rubisco holoenzymes (hereafter referred to as "Rubiscos") and the intrinsically disordered protein EPYC1 combine to phase separate as illustrated in Fig. 1(a). While the structure of Rubisco is strongly constrained by its function in carbon fixation, it is not clear what constraints may in principle apply to EPYC1. In particular, what governs the flexible linkers connecting EPYC1's binding motifs? As shown in Fig. 1(b), we address this question via simulations and theory with a focus on the interplay of EPYC1 linker length and flexibility, strength of binding interactions, and Rubisco sticker location. These microscopic parameters affect how EPYC1 binds to Rubisco. As shown in Fig. 1(c), when EPYC1 is a "good fit," almost all its stickers can bind to a single Rubisco molecule; by contrast for a "poor fit," EPYC1 is not able to bind all its stickers to a single Rubisco. We will explore in detail how this difference influences phase separation.

We begin our study of the two-component Rubisco-EPYC1 system using coarse-grained molecular-dynamics simulations (Fig. 2). Rubisco is modeled according to its crystal structure [38] as a sphere with a diameter of about 12 nm. From cryoelectron microscopy of Rubisco bound to EPYC1 [9], it is known that there are eight stickers on Rubisco for EPYC1 located near 45° from the top and bottom. EPYC1 is modeled as a polymer of five identical stickers based on experiments that show that EPYC1 has five highly similar repeats that bind Rubisco [9]. The effective linker length between EPYC1 stickers is set by a nonlinear spring constant, k, which models the entropic free-energy cost of stretching a linker (see Methods for details). All stickers have a diameter $\sigma = 2$ nm. Sticker-sticker bonds between EPYC1 and Rubisco are implemented via a cosine potential of depth $U_0 = 14\ k_BT$ (see Methods). One-to-one bonding is enforced by excluded volume interactions among all other elements.

Since EPYC1 serves as molecular glue by linking together Rubiscos, it is natural to ask how does EPYC1 linker length

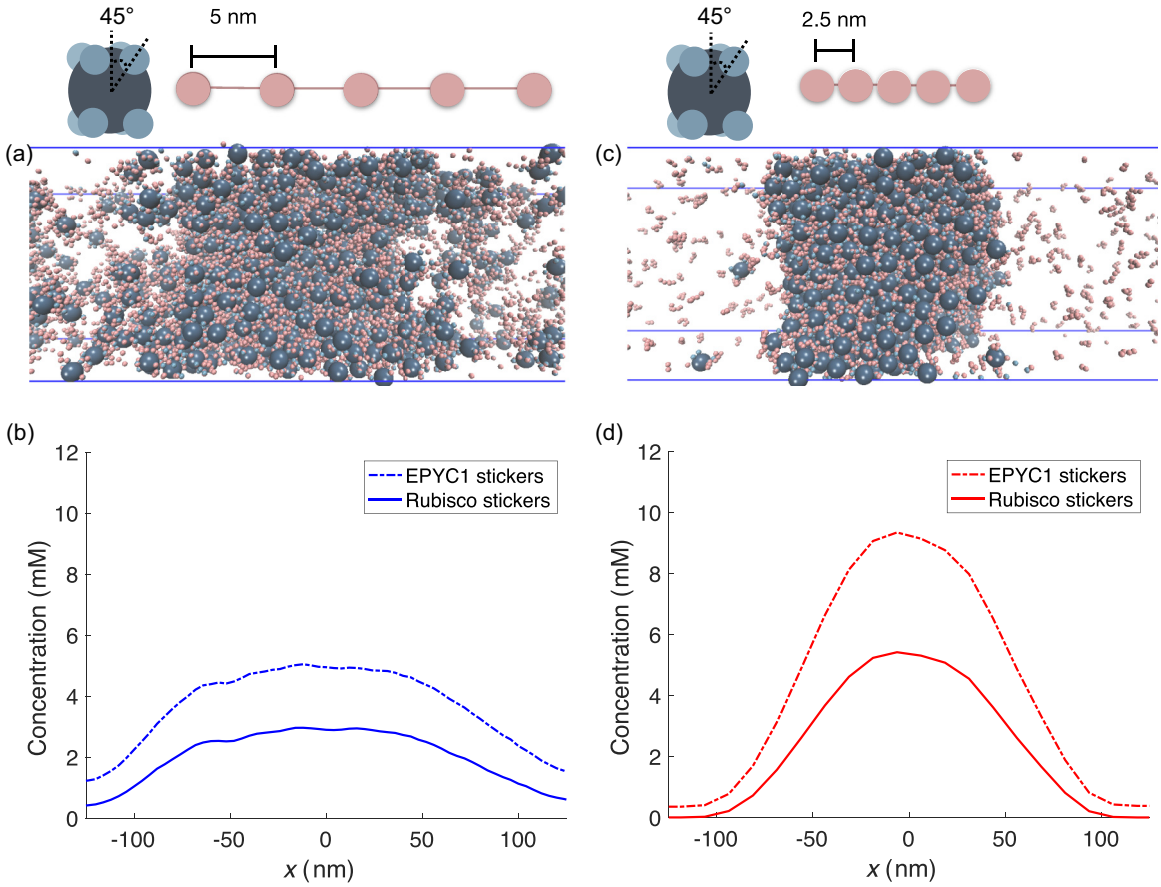

FIG. 2. Rubisco-EPYC1 phase separation is sensitive to EPYC1 linker length. As shown in the schematics, the left and right panels show results for long and short EPYC1 linkers, respectively. In the illustration on the left, the long EPYC1 linker length is 5 nm, while in the illustration on the right, the short EPYC1 linker length is 2.5 nm. In both cases the Rubisco stickers are at 45° from the poles. Data are from simulations of 720 Rubiscos with eight stickers each and 2304 EPYC1s with five stickers each (i.e., twice as many EPYC1 as Rubisco stickers) in a 500 nm × 126 nm × 126 nm box with periodic boundaries; each sticker has a diameter $\sigma = 2$ nm and the sticker binding strength is $U_0 = 14\ k_BT$ (see Methods for details). (a) Snapshot of a phase-separated system with long EPYC1 linkers with a mean equilibrium length of 5 nm ($k = 0.12\ k_BT/\sigma^2$). (b) Sticker concentration profile for EPYC1 (dotted blue) and Rubisco (solid blue) for the system shown in (a) averaged over three simulations. (c) Snapshot of a phase-separated system with short EPYC1 linkers with a mean equilibrium length of 2.5 nm ($k = 1.20\ k_BT/\sigma^2$). (d) Sticker concentration profile for EPYC1 (dotted red) and Rubisco (solid red) stickers for the system shown in (c) averaged over three simulations.

affect phase separation given the fixed locations of Rubisco stickers? To address this question, we first employ a long effective EPYC1 linker length of $l = 5$ nm ($k = 0.12\ k_BT/\sigma^2$) which matches the typical sticker-sticker distance from all atom simulations (see Supplemental Note 1 in the Supplemental Material [39]; also see Refs. [23,40,41]). In Fig. 2, we present results of simulations of Rubisco-EPYC1 systems for these long EPYC1 linkers (left column) using Brownian dynamics at a temperature of $T = 300$ K. In Fig. 2(a), we show a snapshot of a slab of dense-phase condensate surrounded on both sides by the dilute phase, and in Fig. 2(b) we show the corresponding EPYC1 and Rubisco sticker density profiles averaged over three independent runs. The total stoichiometry of 2:1 EPYC1 to Rubisco stickers is near the highest imbalance that still produces phase separation for these long linkers (see Supplemental Note 2 in Ref. [39]).

As shown in Fig. 1(a), the actual binding geometry of EPYC1 and Rubisco requires EPYC1 stickers to be both compact and oriented in a way that shortens the effective linker length between EPYC1 stickers. We therefore compare the above results for long EPYC1 linkers to those for shorter EPYC1 linkers with an effective EPYC1 linker length of $l = 2.5$ nm ($k = 1.20\ k_BT/\sigma^2$). In Fig. 2(c), we show a snapshot for the short-linker system, and in Fig. 2(d) we show the profiles of Rubisco and EPYC1 sticker densities averaged over three independent runs. Comparing the density profiles of the two EPYC1 linker-length systems in Figs. 2(b) and 2(d), we see that short linkers lead to an almost twofold increase of both EPYC1 and Rubisco stickers in the condensed phase. In addition, we observe an approximately tenfold decrease in the critical densities when comparing the long EPYC1 linker system to the short EPYC1 linker system (computed by the ratio of the product of the dilute-phase concentrations of EPYC1s and Rubiscos for the long and short EPYC1 linker systems). Moreover, the short linkers lead to phase separation at much higher stoichiometry differences than the long linkers, for

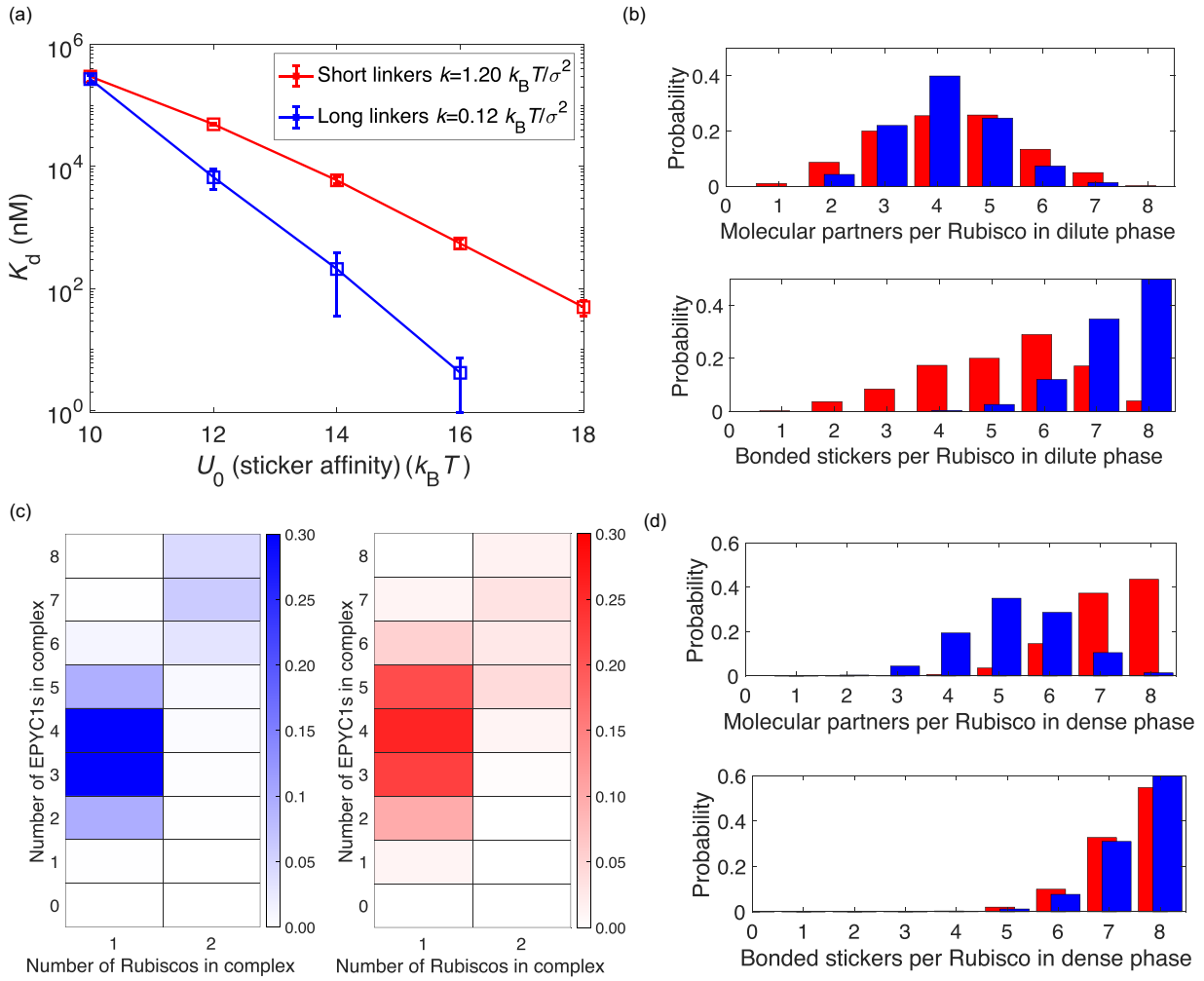

FIG. 3. Simulation results for the dilute- and dense-phase properties of Rubisco-EPYC1 systems with short (red) versus long (blue) EPYC1 linkers with the same parameters as in Fig. 2. (a) Molecular dissociation constant of Rubisco-EPYC1 dimers as a function of sticker affinity (U_0) for both short EPYC1 linkers ($k = 1.20\ k_B T/\sigma^2$) and long EPYC1 linkers ($k = 0.12\ k_B T/\sigma^2$). (b) Upper and lower plots show, respectively, the number of EPYC1 molecular partners bound to each Rubisco and the number of bonded stickers on each Rubisco in the dilute phase. (c) Left and right plots show, respectively, the probability of different complexes in the dilute phase. (d) Upper and lower plots show, respectively, the molecular partners bound to each Rubisco and the number of bonded stickers on each Rubisco in the dense phase.

example, at 10:1 EPYC1 to Rubisco stickers (Supplemental Fig. 2 in Ref. [39]). This implies that the short linkers more strongly favor phase separation than the long linkers under otherwise equivalent conditions.

To better understand the pronounced influence of EPYC1 linker length on phase separation, we next simulate the simple case of a single EPYC1 interacting with a single Rubisco. Specifically, for each of the two different effective linker lengths we obtain the dissociation constant K_d between one EPYC1 and one Rubisco over a range of sticker binding strengths [Fig. 3(a)]. We find that an EPYC1 with long linkers has a lower K_d (stronger binding) to Rubisco since it can form more sticker-sticker bonds with Rubisco than can an EPYC1 with short linkers. As expected, in both cases the dissociation constant K_d decreases exponentially with sticker binding strength. The experimentally estimated dissociation constant between EPYC1 and Rubisco is about 30 nM [11], which corresponds to $U_0 = 14{-}15\ k_B T$ for the long linkers or around $U_0 = 18\ k_B T$ for the short linkers. For comparison, when $U_0 = 14\ k_B T$ the dissociation constant for EPYC1 with short linkers is about 2 mM, i.e., almost two orders of magnitude higher than the K_d for long EPYC1 linkers. A well depth of $U_0 = 14\ k_B T$ does not imply irreversibility since the free energy of binding involves volumetric terms; i.e., $14\ k_B T$ only occurs at a single point. Importantly, the well depth is set to match realistic values of K_d.

The large difference in K_d values for different EPYC1 linker lengths begs the question of how this difference in fit between one EPYC1 and one Rubisco affects the microscale organization of the dilute and dense phases. Considering first the dilute phase, one potentially informative quantity is how many EPYC1 molecules bind to each Rubisco. As shown in Fig. 3(b), we find that these distributions are quite similar for long and short linkers, with a peak around four EPYC1 molecules bound to each Rubisco. However, the number of individual Rubisco stickers bound to EPYC1 stickers is quite

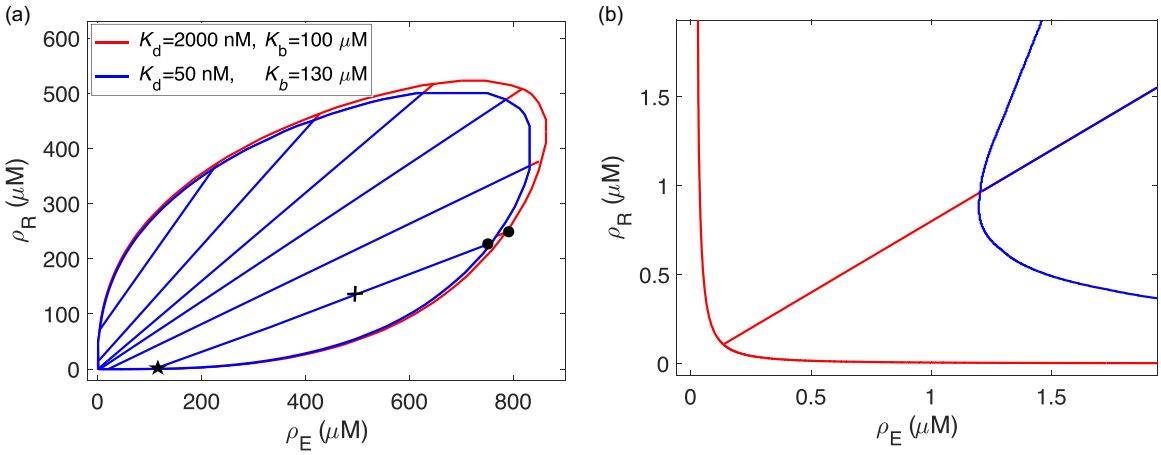

FIG. 4. Dimer-gel theory predictions for EPYC1 linker-length dependence of Rubisco-EPYC1 phase diagram. (a) Predicted phase diagrams for short (red) and long (blue) EPYC1 linkers using the molecular dissociation constants K_d for $U_0 = 14\ k_BT$. The sticker-sticker dissociation constants K_b are fit so that the dilute-phase concentrations for overall concentrations marked with a "+" agree with the corresponding values from the simulations in Fig. 2. ρ_E and ρ_R are the densities of EPYC1 molecules and Rubisco holoenzymes, respectively. Selected tie lines are shown, with the dots indicating dense-phase concentrations for the overall concentrations marked with + (the star denotes two overlapping dilute-phase concentrations). (b) Closeup of the phase diagram in (a) with a representative tie line (see Methods for details).

different: for the long EPYC1 linker system, typically seven or eight Rubisco stickers are bound, whereas for the short-linker system, only four to seven Rubisco stickers are bound. In addition, we ask how Rubisco-EPYC1 association affects the complexes that form in the dilute phase. As shown in Fig. 3(c), most complexes have one Rubisco with about three or four EPYC1s. For the relatively rare complexes with two Rubiscos, there is typically an additional EPYC1 to bridge them. Next, we consider the dense phase. As shown in Fig. 3(d), in sharp contrast to the dilute phase, the typical number of satisfied Rubisco stickers is similarly high about seven or eight for both long and short EPYC1 linkers. However, the number of EPYC1 molecules bound to each Rubisco is quite different, with four to six long-linker EPYC1s versus seven or eight short-linker EPYC1s bound to each Rubisco.

Taken together, these observations suggest an important interplay between EPYC1 linker length and the distance between stickers on a Rubisco. We can gain intuition by considering whether two adjacent EPYC1 stickers are able to bind two adjacent Rubisco stickers. To make this quantitative, we consider a simplified case of a single truncated EPYC1 molecule with only two stickers, one of which is fixed in space, and ask for the probability that the second sticker will be bound to an attractive site a distance d away, assuming an implicit linker given by a simple harmonic spring between the two stickers. This probability is given by (see Methods)

$$p_{\text{bound}} = \left(1 + \frac{\pi^3 l^3 K_b e^{\frac{4d^2}{\pi l^2}}}{8}\right)^{-1}, \tag{1}$$

where l is the effective linker length; K_b is the dissociation constant of a single EPYC1 sticker from the binding site. This result demonstrates that there is a large entropic cost to bind consecutive stickers of EPYC1 when the Rubisco sticker separation represented by d substantially exceeds the effective linker length l between EPYC1 stickers. For a Rubisco sticker spacing of $d = 6.8$ nm, as is the case for Figs. 2 and 3, the exponent $4d^2/(\pi l^2)$ in Eq. (1) implies an entropic cost of only about 1 k_BT for a long linker but a much larger cost of about 10 k_BT for a short linker to extend this distance. Hence, for the short linkers it is too entropically costly for neighboring EPYC1 stickers to bind adjacent Rubisco stickers.

This large difference in entropic costs for long and short linkers to bind neighboring stickers on a Rubisco differentially affects the possible configurations of EPYC1s and Rubiscos in both the dilute and dense phases, which in turn influences phase separation. Specifically, the better fit between the long EPYC1 linkers and the spacing between Rubisco stickers means that a small number of these EPYC1s can fully satisfy a Rubisco, which allows for energetically favorable small complexes of one Rubisco andthree to five EPYC1s that compete with the dense phase. By contrast, such favorable complexes are not available for short linkers—either Rubisco stickers are left unsatisfied as shown in Fig. 3(b) (bottom), or extra EPYC1s would have to be pulled out of solution, with a concomitant translational entropy cost. Instead, the short linkers favor the dense phase within which neighboring EPYC1 stickers can bind to distinct Rubiscos whose stickers can come closer together. Indeed, in the dense phase, short linkers are just as effective as long linkers at satisfying all Rubisco stickers [Fig. 3(d), bottom], albeit with each Rubisco sticker bound by a different EPYC1 molecule [Fig. 3(d), top]. The inverse relationship between molecular fit and phase separation is expanded on further in Supplemental Note 2 in Ref. [39] where we show that when the Rubisco stickers are allowed to move freely on the main spherical base of each Rubisco there is a greatly improved molecular fit for short-linker EPYC1s, which promotes the formation of dilute-phase complexes, and consequently no phase separation is observed at 2:1 EPYC1 to Rubisco sticker stoichiometry. We conclude that the striking differences we observed in our simulations with long versus short EPYC1 linkers can be attributed to the

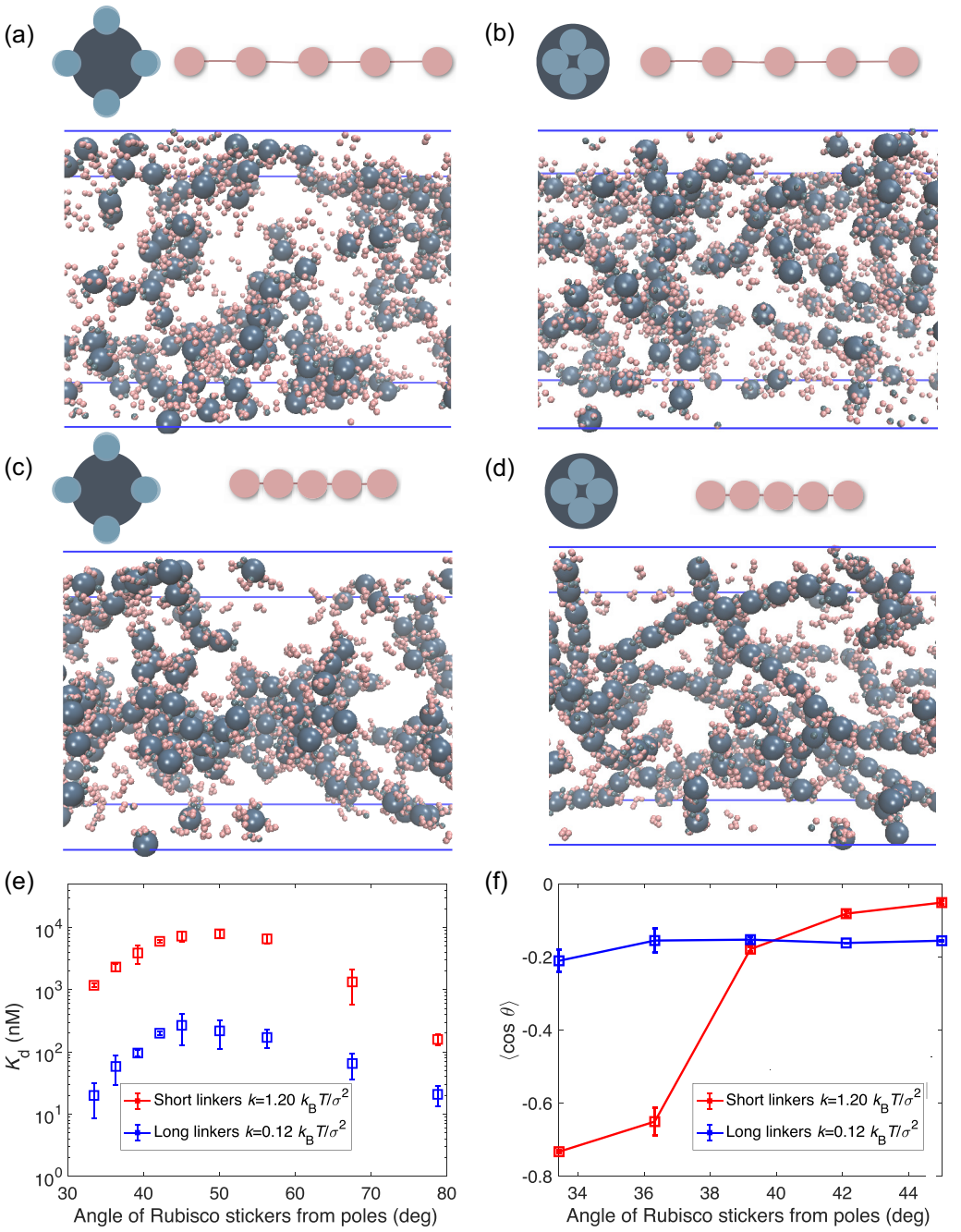

FIG. 5. Rubisco sticker location strongly influences phase separation. Rubisco-EPYC1 systems with long ($k = 0.12\ k_B T/\sigma^2$) and short EPYC1 linkers ($k = 1.20\ k_B T/\sigma^2$) are simulated at 2:1 EPYC1 to Rubisco sticker stoichiometry (608 EPYC1s and 380 Rubiscos) for $U_0 = 14\ k_B T$ in a box of size 315 nm × 126 nm × 126 nm with periodic boundaries at a temperature of $T = 300$ K. All illustrations show a two-dimensional projection of Rubisco from the top. (a), (b) Snapshots of long EPYC1 linkers and Rubiscos whose stickers are located at (a) 79° and (b) 33° from the poles both show weak phase separation. Each system forms a gas of small complexes. (c) Snapshot of short EPYC1 linkers for Rubisco stickers located at 79° from the poles also shows a gas of small complexes. (d) Snapshot of short EPYC1 linkers for Rubisco stickers located at 33° from the poles shows that the system forms a gas of rods. (e) The molecular dissociation constant as a function of Rubisco sticker location for short (red) and long (blue) linkers in a box of size 100 nm × 100 nm × 100 nm. (f) Order parameter $\langle \cos\theta \rangle$ for rods as a function of angle of Rubisco stickers from poles, where θ is the angle formed by neighboring Rubiscos around a central Rubisco (see Methods for details). The blue symbols are for long linkers and the red symbols are for short linkers. Error bars are SD values obtained from three independent simulations.

difference in fit between EPYC1 linker length and Rubisco sticker spacing.

It is natural to ask how this difference in fit at the molecular scale influences the overall phase diagram. To this end, we utilize a minimal dimer-gel theory [11,35] as shown in Fig. 4. While the model is quite simple, it can capture overall behavior independent of many microscopic details, e.g., the exact types and distribution of dilute-phase oligomers,

or dense-phase arrangements. The main components of the model are free-energy densities, including ideal polymer, excluded volume, and binding contributions (see Methods). In the dilute phase, binding is modeled as formation of small oligomers, which for simplicity we limit to Rubisco-EPYC1 dimers. In the dense phase, binding is modeled as a gas of stickers forming independent sticker-sticker bonds. The only binding-dependent inputs to the model are the dissociation constant K_d between one EPYC1 and one Rubisco and the sticker-sticker dissociation constant K_b. For the long and short EPYC1 linkers, the molecular dissociation constants K_d are obtained from Fig. 3(a) for $U_0 = 14\ k_BT$. The sticker-sticker dissociation constants K_b are chosen to match the dilute-phase concentrations found in the simulations in Fig. 2. Figure 4(a) shows the overall phase diagrams obtained for these parameters, with a zoomed-in version of the dilute-phase boundary shown in Fig. 4(b). Consistent with our simulation results, the short EPYC1 linkers strongly favor phase separation, with dilute-phase concentrations as much as an order of magnitude lower than for long linkers. Notably the fitting parameters K_b differ only by 30% compared to the 40-fold difference between the K_d values—thus the striking difference between the phase diagrams can be entirely attributed to the latter. In a nutshell, the better fit between long EPYC1 linkers and adjacent Rubisco stickers allows for stable small complexes in the dilute phase, which disfavors formation of a condensate.

Up to this point, we have changed the effective linker length of EPYC1 while keeping the Rubisco stickers at their fixed native locations at 45° from the poles. However, one may ask why evolution has placed the Rubisco stickers of *C. reinhardtii* at particular positions. In Fig. 5, we study the importance of Rubisco sticker location on phase separation and molecular fit. We initially consider the long EPYC1 linkers with mean length 5 nm ($k = 0.12\ k_BT/\sigma^2$). Moving the Rubisco stickers either farther away from the poles to 79° [Fig. 5(a)] or closer to the poles to 33° [Fig. 5(b)] both result in weaker phase separation for long linkers. These results are consistent with our above observations that stronger affinity between EPYC1 and Rubisco (lower molecular K_d), suppresses phase separation by favoring dimers or other small oligomers of EPYC1 and Rubisco over the condensed phase; as seen in Fig. 5(e) the K_d values for long linkers are largest near the native Rubisco sticker location at 45°, and decrease as the stickers are moved either farther from or closer to the poles of Rubisco. Thus, the naturally occurring Rubisco sticker spacing produces the worst molecular fit, but is thereby optimal for phase separation. (Note that changing the sticker-sticker affinity, U_0, has little effect on phase separation because, in the language of the dimer-gel theory, it changes both K_d and K_b together, leaving the phase diagram qualitatively unchanged.)

Motivated by these results, we wonder whether moving the Rubisco stickers closer to the poles, and thus closer together, would improve the molecular fit between Rubisco and a single EPYC1 with short linkers of the mean length 2.5 nm ($k = 1.20\ k_BT/\sigma^2$), thus lowering K_d, and disfavoring phase separation. In Fig. 5(e), one sees that the molecular fit does improve for short linkers when the Rubisco stickers move closer to the poles. While moving the Rubisco stickers all the way to 33° did indeed abrogate phase separation [Fig. 5(d)], we are surprised to see that the system organized into a gas of Rubisco "rods," which does not occur in the long EPYC1 linker system. In Fig. 5(f), we quantify this unique organization using as a simple order parameter $\langle \cos\theta \rangle$, i.e., the average cosine of the angle formed by all pairs of neighboring Rubiscos about each central one, averaged over all Rubiscos (see Methods). For the phase-separated state of short-linker EPYC1s and Rubiscos with stickers at 45°, we find $\langle \cos\theta \rangle$ approximately equals 0, consistent with random packing of Rubiscos; in contrast, for the gas of rods that forms for Rubisco stickers at 33°, $\langle \cos\theta \rangle$ approximately equals –0.8, implying close to complete rodlike alignment. The reason for rod formation appears to be quite simple: moving the Rubisco stickers to 33° means that two Rubiscos stacked pole to pole can have four extremely close pairs of stickers, which can be easily bridged by EPYC1 linkers.

One interesting question is, what is the molecular fit that the system would "prefer" to have? We address this question via a toy model in which the Rubisco stickers are allowed to diffuse freely around the Rubisco sphere (Supplemental Fig. 4 in Ref. [39]). We find that Rubisco stickers group together leading to a much higher affinity for short-linker EPYC1 (lower K_d), favoring formation of a gas of small oligomers instead of phase separation, but there is no tendency to form rods. This further confirms the conclusion that molecular fit is inversely related to phase separation.

III. DISCUSSION

Biomolecular condensates are typically "network liquids" held together by specific interactions between sticker domains connected by flexible linkers. While linkers play multiple known roles including occupying volume [23,24], adding attractive interactions [15–17], and recruiting clients [21,22], the relation between linker length and other intrinsic length scales has been less studied. We explore this topic in the context of a model for the algal pyrenoid, a condensate formed by specific interactions between the intrinsically disordered protein EPYC1 and the rigid holoenzyme Rubisco. Using coarse-grained molecular-dynamics simulations, we find that EPYC1s with shorter linkers led to substantially stronger phase separation. We trace this effect to a worse fit between the shortened EPYC1 linkers and the spacing between stickers on a Rubisco: adjacent stickers on short-linker EPYC1s cannot stretch far enough to bind adjacent stickers on Rubisco. By contrast, in the dense phase, adjacent stickers on EPYC1 can bind to distinct neighboring Rubiscos; consequently, short EPYC1 linkers favor the dense phase. Using a minimal dimer-gel theory to predict the full phase diagram, we find that short linkers can lead to phase separation at up to an order of magnitude lower concentrations than longer linkers. Additionally, by moving Rubisco sticker locations closer to the poles, we find that the native Rubisco sticker spacing is optimal for phase separation, and discover an unusual state in which Rubiscos form a gas of rods. Our study thus highlights how the interplay of linker length with other relevant length scales, in this case Rubisco sticker spacing, can dramatically influence biomolecular phase separation.

We model EPYC1 linkers implicitly with an effective length set by a spring constant. This allows us to isolate

the effect of EPYC1 linker length on phase separation. The real Rubisco-EPYC1 system, even *in vitro*, is certainly more complicated than our simple model: linkers occupy volume and may engage in attractive interactions, while the amino-acid sequences of EPYC1 stickers include charged residues and are not identical, and these stickers bind directionally to Rubisco. However, we do not expect these details to affect our overall conclusion that the phase diagram and the microscale organization of the dense and dilute phases depends sensitively on the fit between EPYC1 linkers and Rubisco stickers. Going beyond the current model would nevertheless be valuable in understanding the interplay of the realistic factors noted above. For example, explicit modeling of linkers might reveal if the linker length is tuned to a happy medium—long enough to bridge Rubisco stickers but not so long as to interfere with phase separation by excluded volume. Improvements to make the dimer-gel theory more quantitative are possible as well—our simulations indicate that higher oligomers can dominate over dimers in the dilute phase, and correlations between sticker-sticker bonds in the dense phase may effectively renormalize the density of independent stickers [35].

The observed sensitive dependence of phase separation on EPYC1 linker length and Rubisco sticker spacing raises the question whether these two lengths may have been jointly optimized by evolution. In the photosynthetic alga *C. reinhardtii*, the pyrenoid is required for efficient carbon capture and so is under strong functional selection. Interestingly, the pyrenoid undergoes complex dynamics during the cell cycle, dissolving into the chloroplast stroma prior to cell division and re-forming in the two daughter cells. Hence, evolutionary optimization may not simply favor strong phase separation, but rather may allow for rapid transitions between a phase-separated and a dissolved state, e.g., upon phosphorylation of EPYC1. Moreover, it is essential to the pyrenoid's function that it remain liquid, as a relatively small number of Rubisco activase proteins must be able to move freely within the pyrenoid matrix to remove inhibitory substrates from Rubisco catalytic sites. Given these multiple constraints, we are hesitant to make concrete statements about the "best" length of the EPYC1 linker, but we will state what functional requirements the real linker would need to satisfy. The illustration in Fig. 1(a) demonstrates one way in which Rubisco can form the pyrenoid matrix as observed from cryoelectron tomography [9,10]. One observation is that an EPYC1 linker would need to bind consecutive stickers of the same Rubisco holoenzyme with proper directionality of binding. In Ref. [9] it was found that within the condensate, the distance between the closest EPYC1 stickers of proximal Rubiscos is about 4 nm, which is close to the long-linker length. However, the Rubisco-EPYC1 interaction is not completely understood, especially considering that residues of EPYC1 are known to be phosphorylated under some conditions [42]. It could be that phosphorylation changes the effective linker length by turning some EPYC1 stickers off, or more generally, by modifying the Rubisco-EPYC1 interaction. Another functional requirement is that Rubisco-EPYC1 condensates are liquidlike with a high degree of mixing as observed in fluorescence recovery after photobleaching (FRAP) experiments [10]. Thus, the EPYC1 linker lengths must be long enough to allow some space between Rubiscos to permit the flow of other pyrenoid-localized proteins.

The concept of linker length is not only of interest in the context of the primary condensate-forming linker, EPYC1, but also in the context of other proteins that localize to the condensate and possess multiple Rubisco-binding motifs like those of EPYC1. Specific examples in *C. reinhardtii* are Rubisco-binding membrane proteins (RBMPs) which are localized to the membrane tubules that supply CO_2 to Rubisco, and StArch Granules Abnormal (SAGA) proteins which are localized to the starch sheath [8]. Both protein families have multiple Rubisco-binding motifs, spaced in some cases by linkers of similar length to those of EPYC1 [about 70 amino acids (AAs)] or about half that length (about 33 AAs). Having sets of longer and shorter linkers may have functional purposes. One possibility is that the longer linkers may permit binding to a single Rubisco, while the shorter linkers may promote bridging between two Rubiscos, or the longer lengths may still favor bridging Rubiscos, but over larger distances. The tendency to bridge Rubiscos would favor localization of RBMP and SAGA proteins with the Rubisco-EPYC1 condensate where many Rubiscos are in close proximity, rather than to Rubiscos dissolved throughout the stroma. Moreover, bridging configurations would leave open stickers on each Rubisco for EPYC1, allowing the bridged Rubiscos to remain part of the condensate, rather than RBMPs and SAGAs sequestering Rubiscos away from the pyrenoid.

The concept of molecular fit may also apply to linker-Rubisco interactions within alpha [13] and beta [14] cyanobacteria. Each Rubisco holoenzyme in alpha-carboxysomes has eight stickers for its flexible partner protein, CsoS2, grouped into four north-south pairs evenly spread around the equator. In the alpha cyanobacteria *Halothiobacillus neapolitanus*, CsoS2 has four stickers [13] for Rubisco on its N-terminal domain with an approximate linker length (about 50 AAs) of 6.8 nm (see Methods). By contrast, each Rubisco in beta-carboxysomes has only four stickers for its flexible partner CcmM, and these are evenly spread around the equator. In the beta cyanobacteria *Synechococcus elongatus*, CcmM has three stickers [14] for Rubisco separated by short linkers (about 30 AAs) with an approximate linker length of 5.2 nm (see Methods). Since the diameter of Rubisco is about 12 nm, both the north-south pairs of Rubisco stickers in alpha-carboxysomes and the individual stickers in beta-carboxysomes are separated by about 9.4 nm around the equator (see Methods). This implies that, like the Rubisco-EPYC1 pyrenoid system, the geometry of binding may favor condensates over small oligomers in carboxysomes. Namely, the stickers on carboxysome Rubiscos may be too far apart for a single partner protein, CsoS2 or CcmM, to fully bond to one Rubisco, favoring instead bridging of multiple Rubiscos and so driving phase separation.

The kind of interplay between length scales we explored in the Rubisco-EPYC1 system may be expected as a general feature of multidomain proteins containing intrinsically disordered linkers. Specifically, Sørensen *et al.* [43] considered two proteins, each with two stickers connected by flexible linkers such that the proteins could form bivalently bound pairs. They found that avidity, i.e., the enhancement of overall binding due to multivalency, increased with decreasing linker length,

presumably due to the increase in the effective concentration of the potential binding partners, albeit with a scaling with linker length that suggested additional linker-associated interactions. This interplay of effective linker length with binding constants in multivalent biomolecules is quite ubiquitous [44] and can be viewed in terms of an effective concentration of stickers set by linker length [45,46] relative to sticker affinity. Moreover, the functional importance of the linker length between binding domains is attested to by the conservation of actual [47] or effective [48] linker length.

We show via simulations and theory that effective linker length relative to another intrinsic length scale can have a drastic effect on phase separation. In the context of Rubisco-EPYC1 condensation, our predictions could be tested by constructing EPYC1 linkers of differing lengths and measuring changes in the phase diagram, as well as probing for microscale features such as Rubisco rods. In live cells, condensates typically form transiently, e.g., in *C. reinhardtii* the pyrenoid dissolves and re-forms during cell division. Thus, condensate dynamics is also important for cellular function, and it is known that linker properties can affect dynamical properties such as diffusion constants and viscosity [18,19]. We hope our results will inspire future investigations of how cells may use linker properties to tune both the steady-state and dynamical properties of biomolecular condensates.

IV. METHODS

A. Model details

The LAMMPS package [49] is used for simulations. In the model, the rigid molecule, Rubisco, has a spherical base with a diameter of 11.6 nm and has eight stickers of diameter $\sigma = 2$ nm at 45° from the top and bottom and uniformly distributed around the circumference (as shown schematically in Fig. 2). The flexible molecule, EPYC1, is modeled as a linear polymer with five stickers of diameter $\sigma = 2$ nm connected by implicit linkers described by a finite extensible nonlinear elastic (FENE) potential [50]:

$$U_{\rm b}(r) = -\frac{1}{2}kR_0^2 \log\left[1 - \left(\frac{r}{R_0}\right)^2\right]. \tag{2}$$

Here, $R_0 = 19$ nm is the length of the linker when fully extended, r is the relative distance between two consecutive EPYC1 stickers, and k is the spring constant which tunes the effective linker length. We choose an extended length of $R_0 = 19$ nm since an amino acid residue is 0.3–0.4 nm long [51], so the extended length of the 60 amino acid EPYC1 linkers is about 18–24 nm. Our value for $R_0 = 19$ nm is thus within the range of the estimated linker length for this system. The sticker-sticker association between EPYC1 and Rubisco is given by the following potential:

$$U_{\rm a}(r) = -\frac{U_0}{2}\left[1 + \cos\left(\frac{\pi r}{r_{\rm c}}\right)\right], \quad r < r_{\rm c}, \tag{3}$$

where U_0 and the cutoff distance $r_{\rm c} = \sigma/2 = 1$ nm define the magnitude and range of the attraction. Additionally, there are excluded volume interactions between all elements of EPYC1 and Rubisco (except stickers of opposite type) given by the potential:

$$U_{\rm r}(r) = 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} + \frac{1}{4}\right], \quad r \leqslant 2^{1/6}\,\sigma_{ij}, \tag{4}$$

where $\varepsilon = 1\ k_{\rm B}T$, σ_{ij} is the effective diameter, and i and j denote the interacting components. For interactions between two EPYC1 stickers or between two Rubisco stickers the effective diameter is $\sigma_{\rm EE} = \sigma_{\rm RR} = \sigma = 2$ nm, while for interactions between the Rubisco base and either a Rubisco sticker or an EPYC1 sticker the effective diameter is $\sigma_{\rm RE} = 13.6$ nm. The attractive cosine potential between Rubisco and EPYC1 stickers [Eq. (3)], along with the strong excluded volume interactions between all other elements [Eq. (4)], guarantees at most one-to-one bonding between the Rubisco and EPYC1 stickers.

B. Dissociation constant $K_{\rm d}$

For Fig. 2(a) of the main text, the molecular dissociation constant is measured in a box of 50 nm × 50 nm × 50 nm with periodic boundaries containing a single EPYC1 and a single Rubisco over a time of 10 ns with a time step of 1 ps. The EPYC1 and Rubisco are considered to be bound when at least one EPYC1 sticker is within the attractive-interaction cutoff distance $r_{\rm c}$ of a Rubisco sticker. From two-state kinetics, with one state being the bound state and the other being the unbound state, the dissociation constant is obtained from the following equation:

$$P_{\rm b} = \frac{[\text{EPYC1}]}{K_{\rm d} + [\text{EPYC1}]}, \tag{5}$$

where $P_{\rm b}$ is the probability, or fraction of time, the two molecules are bound, and the concentration of EPYC1 is given by

$$[\text{EPYC1}] \approx \frac{1}{N_{\rm A}V}, \tag{6}$$

where $N_{\rm A}$ is Avogadro's number and V is the volume of the simulation box in liters.

C. Determination of the dense and dilute phases, satisfied bonds, molecular partners, and complexes

In Fig. 3 of the main text, properties of the dense phase and the dilute phase are considered separately. This requires distinguishing the two phases. The dense phase and dilute phase are determined by density analysis. To this end, a density profile was measured every 10 ns from snapshots such as those shown in Figs. 2(a) and 2(b). For each time point, the density profile is shifted to put the center of mass of the Rubisco molecules at the center of the box, which centers the dense phase. Subsequently, the "dense phase" is defined as the central region away from interface with the dilute phase—in practice, we choose the region where the average Rubisco density is at least 80% of its maximum. Similarly, the "dilute phase" is defined as the flanking regions away from the interface. Because of the larger region of dilute phase, we include only zones well away from the interface, where the average density is flat within numerical noise.

To enumerate the complexes found in the dilute phase as shown in Fig. 3(c), we employ a custom cluster code [52]. The cluster code computes molecular bonds between Rubiscos and EPYC1s via the bonding criteria described in the section titled Dissociation constant K_d.

From the same definition of a bond, the number of satisfied stickers is enumerated for each Rubisco in both the dense and dilute phases, along with the number of EPYC1 molecular partners per Rubisco, as shown in Figs. 3(b) and 3(d).

D. Dimer-gel theory

We consider a dimer-gel theory which has three terms (see Refs. [11,35]):

$$F = F_{\rm ni} + F_{\rm ex} + F_{\rm s}. \tag{7}$$

The first term considers noninteracting, ideal, contributions to the free-energy density given by

$$F_{\rm ni} = \frac{c_{\rm R}}{L_{\rm R}} \ln \frac{c_{\rm R}}{e\, L_{\rm R}} + \frac{c_{\rm E}}{L_{\rm E}} \ln \frac{c_{\rm E}}{e\, L_{\rm E}}, \tag{8}$$

where $c_{\rm R}$ and $c_{\rm E}$ are the sticker concentrations for Rubisco and EPYC1, and $L_{\rm R} = 8$ and $L_{\rm E} = 5$ are the number of stickers for a single Rubisco and single EPYC1, respectively. The logarithms in Eq. (8) are dimensionless with a reference fluid concentration set to unity. The second term $F_{\rm ex}$ accounts for excluded volume interactions among the molecules, which we take to be hard spheres:

$$F_{\rm ex} = \nu_{\rm R} c_{\rm R}^2 + \nu_{\rm E} c_{\rm E}^2 + \nu_{\rm ER} c_{\rm E} c_{\rm R}, \tag{9}$$

where the effective volume constants $\nu_{\rm R}$, $\nu_{\rm E}$, and $\nu_{\rm ER}$ are obtained by a virial expansion [53,54]. The Rubisco-Rubisco effective volume is four times an eighth of its molecular volume with diameter $d_{\rm R} = 10$ nm given by $\nu_{\rm R} = 4\,\frac{4}{3}\frac{\pi}{8}(\frac{d_{\rm R}}{2})^3 = 261.8$ nm^3. The EPYC1-EPYC1 effective volume is four times its effective sticker volume, where a sticker for EPYC1 consists of approximately 60 amino acids that are each 3−4 Å in length, which includes the length of the region responsible for binding Rubisco as well as a linker [51], in terms of its radius of gyration $R_{\rm g} \approx 1$ nm given by $\nu_{\rm E} = 4\,\frac{4}{3}\pi R_{\rm g}^3 = 16.76$ nm^3. The effective Rubisco-EPYC1 volume is determined by the effective radius of the two such that $\nu_{\rm ER} = 8\frac{4}{3}\pi(\frac{d_R}{8} + \frac{R_{\rm g}}{2})^3 = 179.60$ nm^3. These volumes are written in terms of molarity which involves multiplication by Avogadro's number after being converted to liters.

The final term, $F_{\rm s}$, which describes sticker-sticker bonding, is taken to be the minimum of the free-energy density to form molecular dimers or the free-energy density to form independent sticker-sticker pairs (since we find one of these free energies always dominates the other),

$$F_{\rm s} = \min\left(F_{\rm dim}, F_{\rm ind}\right), \tag{10}$$

with

$$F_{\rm dim} = \rho_{\rm d} \ln K_{\rm d} + \rho_{\rm d} \ln \frac{\rho_{\rm d}}{e} + (\rho_{\rm R} - \rho_{\rm d}) \ln \frac{(\rho_{\rm R} - \rho_{\rm d})}{e} + (\rho_{\rm E} - \rho_{\rm d}) \ln \frac{(\rho_{\rm E} - \rho_{\rm d})}{e} - \rho_{\rm R} \ln \frac{\rho_{\rm R}}{e} - \rho_{\rm E} \ln \frac{\rho_{\rm E}}{e}, \tag{11}$$

and

$$F_{\rm ind} = c_{\rm b} \ln K_{\rm b} + c_{\rm b} \ln \frac{c_{\rm b}}{e} + (c_{\rm R} - c_{\rm b}) \ln \frac{(c_{\rm R} - c_{\rm b})}{e} + (c_{\rm E} - c_{\rm b}) \ln \frac{(c_{\rm E} - c_{\rm b})}{e} - c_{\rm R} \ln \frac{c_{\rm R}}{e} - c_{\rm E} \ln \frac{c_{\rm E}}{e}. \tag{12}$$

Here, $\rho_{\rm R}$ and $\rho_{\rm E}$ are the molecular concentrations of Rubisco and EPYC1, $c_{\rm R}$ and $c_{\rm E}$ are the respective sticker concentrations of Rubisco and EPYC1, and $\rho_{\rm d}$ and $c_{\rm b}$ are the concentrations of molecular dimers and independent sticker-sticker pairs given, respectively, by the following:

$$\rho_{\rm d} = \tfrac{1}{2}[\rho_{\rm R} + \rho_{\rm E} + K_{\rm d} - \sqrt{(\rho_{\rm R} + \rho_{\rm E} + K_{\rm d})^2 - 4\rho_{\rm E}\rho_{\rm R}}], \tag{13}$$

and

$$c_{\rm b} = \tfrac{1}{2}[c_{\rm R} + c_{\rm E} + K_{\rm b} - \sqrt{(c_{\rm R} + c_{\rm E} + K_{\rm b})^2 - 4c_{\rm E}c_{\rm R}}]. \tag{14}$$

The logarithms in Eqs. (11) and (12) are dimensionless with a reference fluid concentration set to unity. The convex hull of the total free-energy density F determines the phase diagram. In practice, the only additional inputs required by the dimer-gel theory are the molecular dissociation constant, $K_{\rm d}$, and the sticker-sticker dissociation constant, $K_{\rm b}$. In Fig. 4, we use the dissociation constant $K_{\rm d}$ from Fig. 3(a) that corresponds to $U_0 = 14\ k_{\rm B}T$ for each of the two linker lengths. The $K_{\rm b}$ value is fit separately for each linker length so that the dilute-phase concentrations within the dimer-gel theory match the simulation results shown in Fig. 2.

E. Rodlike order parameter

The transition to a gas of rods is quantified by the alignment of Rubisco molecules given by the average cosine of the relative angle between all pairs of Rubisco that are neighbors of the ith Rubisco,

$$\langle \cos\theta_i \rangle = \frac{1}{n_i} \sum_{j,k \in \text{cutoff}} \frac{\boldsymbol{r_{ij}} \cdot \boldsymbol{r_{ik}}}{r_{ij} r_{ik}}, \tag{15}$$

where bold $\boldsymbol{r_{ij}}$ and $\boldsymbol{r_{ik}}$ are, respectively, the vectors from the center of Rubisco i to the centers of Rubiscos j and k, for all neighbors whose absolute distances r_{ij} and r_{ik} fall within a cutoff distance of 15 nm to include only nearest neighbors. n_i is the number of Rubisco pairs enumerated in the sum. This observable is then averaged over all Rubisco molecules to get

$$\langle \cos\theta \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle \cos\theta_i \rangle. \tag{16}$$

F. Conditional probability to bond two consecutive stickers in a two-sticker system

To understand why bonded configurations of EPYC1 and Rubisco depend on EPYC1 linker length, we first consider a simpler system of a truncated EPYC1 molecule (EPYC1_2) consisting of two stickers connected by a linker. There is no excluded volume between the stickers and the linker is modeled as a harmonic spring with energy

$$U_{\rm h}(r) = \tfrac{1}{2}\kappa r^2, \tag{17}$$

where r is the relative distance between the two stickers, and κ is the spring constant. For this system, the equilibrium

effective linker length is given by

$$l = \frac{1}{Z_{\text{free}}} \int r e^{-\beta U_{\text{h}}(r)} d^3 r = \sqrt{\frac{8}{\pi \beta \kappa}}\,, \tag{18}$$

where $\beta = \frac{1}{k_{\text{B}} T}$, and

$$Z_{\text{free}} = \int e^{-\beta U_{\text{h}}(r)} d^3 r = \sqrt{\frac{8\,\pi^3}{\beta^3 \kappa^3}}\,. \tag{19}$$

From Eq. (18), we can relate the spring constant to the effective linker length,

$$\kappa = \frac{8}{\pi \beta l^2}. \tag{20}$$

Now consider two stickers for the EPYC1$_2$ stickers that are a distance d apart, with one of the stickers at $\boldsymbol{r}_1 = (0, 0, 0)$ and the other at $\boldsymbol{r}_2 = (d, 0, 0)$. Given that one sticker of EPYC1$_2$ is bound at exactly $\boldsymbol{r}_1$, we want to calculate the conditional probability that the other sticker is bound at $\boldsymbol{r}_2$. At thermal equilibrium, this conditional probability is given by

$$p_{\text{bound}} = \frac{Z_{\text{bound}}}{Z_{\text{free}} + Z_{\text{bound}}}, \tag{21}$$

where Z_{bound} is the partition function of the second sticker being bound at a distance, d,

$$Z_{\text{bound}} = v e^{-\beta U_{\text{h}}(d)} e^{+\beta U_0}, \tag{22}$$

where v is the volume over which a bond can be made, and U_0 is the bonding energy. More generally, we can express Z_{bound} in terms of the sticker dissociation constant, $K_{\text{b}} = \frac{e^{-\beta U_0}}{v}$, so that

$$Z_{\text{bound}} = \frac{e^{-\beta U_{\text{h}}(d)}}{K_{\text{b}}}. \tag{23}$$

Thus, the conditional probability for the second sticker to be bound is

$$p_{\text{bound}} = \left(1 + \frac{\pi^3 l^3 K_{\text{b}} e^{\frac{4d^2}{\pi l^2}}}{8}\right)^{-1}. \tag{24}$$

The result in Eq. (24) implies that it is unfavorable for two adjacent EPYC1 stickers to bind neighboring Rubisco stickers if the effective EPYC1 linker length l is substantially smaller than the sticker spacing d. This has strong implications for the observation of Rubisco-EPYC1 bonding arrangements in both dense and dilute phases in our simulations.

G. Estimation of linker lengths and Rubisco sticker spacing in carboxysomes

Approximating a disordered protein as an ideal chain, the radius of gyration is given by $R_{\text{g}} = \sqrt{N} l_{\text{AA}}$, where N is the number of amino acids in the protein and $l_{\text{AA}} \approx 0.39$ nm is the length of an amino acid [51]. For a protein composed of stickers and linkers, the effective linker length is then given by $l = \sqrt{6} R_{\text{g}}$ in terms of a single linker's radius of gyration (see Supplemental Note 1 in Ref. [39]). If we approximate CsoS2 linkers as 50 AAs, this gives a radius of gyration of $R_{\text{g}} = 2.8$ nm and an effective linker length of $l = 6.8$ nm. Likewise, CcmM linkers are about 30 AAs, which gives a radius of gyration of $R_{\text{g}} = 2.1$ nm and an effective linker length of $l = 5.2$ nm.

Next, we would like to estimate the distance between Rubisco stickers for alpha and beta cyanobacteria. Given that Rubisco has a diameter of 12 nm, its circumference is $C = 2\pi$ (6 nm), about 37.7 nm. For alpha cyanobacteria, the spacing between adjacent sets of north-south sticker pairs is therefore $C/4 = 9.4$ nm, and the distance between the individual Rubisco stickers in beta cyanobacteria is the same.

Data or codes supporting the findings of this paper are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

We thank Ariel Amir, Guanhua He, Shan He, Wencheng Ji, Ofer Kimchi, Linnea Lemma, Sam Safran, David Sivak, and Jonathan Yuly for useful discussions. This work was supported by grants from the National Institutes of Health (Grant No. R01GM140032), the National Science Foundation (Grant No. MCB-1935444), and the National Science Foundation through the Center for the Physics of Biological Function (Grant No. PHY-1734030). T.G. was supported by the Schmidt Science Fellowship.

T.G., Y.Z, M.C.J., and N.S.W. designed the study. T.G. performed all simulations and analytical calculations. T.G. and N.S.W. developed theory. All authors contributed to interpreting results and preparing the paper.

The authors declare no competing interests.

[1] X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, and R. D. Vale, Phase separation of signaling molecules promotes T cell receptor signal transduction, Science **352**, 595 (2016).

[2] S. Sun, T. GrandPre, D. T. Limmer, and J. T. Groves, Kinetic frustration by limited bond availability controls the LAT protein condensation phase transition on membranes, Sci. Adv. **8**, eabo5295 (2022).

[3] L. B. Case, X. Zhang, J. A. Ditlev, and M. K. Rosen, Stoichiometry controls activity of phase-separated clusters of actin signaling proteins, Science **363**, 1093 (2019).

[4] B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga *et al.*, Coactivator condensation at super-enhancers links phase separation and gene control, Science **361**, eaar3958 (2018).

[5] M. Mir, A. Reimer, J. E. Haines, X. Li, M. Stadler, H. Garcia, M. B. Eisen, and X. Darzacq, Dense Bicoid hubs accentuate binding along the morphogen gradient, Genes Dev. **31**, 1784 (2017).

[6] S. An, R. Kumar, E. D. Sheets, and S. J. Benkovic, Reversible compartmentalization of *de novo* purine biosynthetic complexes in living cells, Science **320**, 103 (2008).

[7] W. S. L. Ang, J. A. How, J. B. How, and O. Mueller-Cajar, The stickers and spacers of Rubiscondensation: Assembling the centrepiece of biophysical CO_2-concentrating mechanisms, J. Exp. Bot. **74**, 612 (2023).

[8] M. T. Meyer, A. K. Itakura, W. Patina, L. Wang, S. He, T. Emrich-Mills, C. S. Lau, G. Yates, L. C. M. Mackinder, and M. C. Jonikas, Assembly of the algal CO_2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif, Sci. Adv. **6**, eabd2408 (2020).

[9] S. He, H. T. Chou, D. Matthies, T. Wunder, M. T. Meyer, N. Atkinson, A. Martinez-Sanchez, P. D. Jeffrey, S. A. Port, W. Patena *et al.*, The structural basis of Rubisco phase separation in the pyrenoid, Nat. Plants **6**, 1480 (2020).

[10] E. S. Rosenzweig, B. Xu, L. K. Cuellar, A. Martinez-Sanchez, M. Schaffer, M. Strauss, H. N. Cartwright, P. Ronceray, J. M. Plitzko, F. Forster *et al.*, The eukaryotic CO_2-concentrating organelle is liquid-like and exhibits dynamic reorganization, Cell **171**, 148 (2017).

[11] G. He, T. GrandPre, H. Wilson, Y. Zhang, M. C. Jonikas, N. S. Wingreen, and Q. Wang, Phase-separating pyrenoid proteins form complexes in the dilute phase, Commun. Biol. **6**, 19 (2023).

[12] J. Barrett, P. Girr, and L. C. M. Mackinder, Pyrenoids: CO_2-fixing phase separated liquid organelles, Biochim. Biophys. Acta. Mol. Cell Res. **1868**, 118949 (2021).

[13] L. M. Oltrogge, T. Chaijarasphong, A. W. Chen, E. R. Bolin, S. Marqusee, and D. F. Savage, Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation, Nat. Struct. Mol. Biol. **27**, 281 (2020).

[14] H. Wang, X. Yan, H. Aigner, A. Bracher, N. D. Nguyen, W. Y. Hee, B. M. Long, G. D. Price, F. U. Hartl, and M. Hayer-Hartl, Rubisco condensate formation by CcmM in β-carboxysome biogenesis, Nature (London) **566**, 131 (2019).

[15] Y. Wang, H. Zhou, X. Sun, Q. Huang, S. Li, Z. Liu, C. Zhang, and L. Lai, Charge segregation in the intrinsically disordered region governs VRN1 and DNA liquid-like phase separation robustness, J. Mol. Biol. **433**, 167269 (2021).

[16] S. Banjade, Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu, and M. K. Rosen, Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck, Proc. Natl. Acad. Sci. USA **112**, E6426 (2015).

[17] B. S. Schuster, G. L. Dignon, W. S. Tang, F. M. Kelley, A. K. Ranganath, C. N. Janke, A. G. Simpkins, R. M. Regy, D. A. Hammer, M. C. Good *et al.*, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior, Proc. Natl. Acad. Sci. USA **117**, 11421 (2020).

[18] R. S. Fisher and S. Elbaum-Garfinkle, Tunable multiphase dynamics of arginine and lysine liquid condensates, Nat. Commun. **11**, 4628 (2020).

[19] J. D. Schmit, F. He, S. Mishra, R. R. Ketchem, C. E. Woods, and B. A. Kerwin, Entanglement model of antibody viscosity, J. Phys. Chem. B **118**, 5044 (2014).

[20] I. Alshareedahet, M. M. Moosa, M. Pham, D. A. Potoyan, and P. R. Banerjee, Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides, Nat. Commun. **12**, 6620 (2021).

[21] K. Kasahara, M. Shiina, J. Higo, K. Ogata, and H. Nakamura, Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an auto-inhibitory state, Nucl. Acids Res. **46**, 2243 (2018).

[22] B. S. Schuster, M. C. Good, and D. A. Hammer, Controllable protein phase separation and modular recruitment to form responsive membraneless organelles, Nat. Commun. **9**, 2985 (2018).

[23] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, Oxford, UK, 2003).

[24] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins, eLife **6**, e30294 (2017).

[25] S. Ranganathan and E. I. Shakhnovich, Dynamic metastable long-living droplets formed by sticker-spacer proteins, eLife **9**, e56159 (2020).

[26] J. McCarty, K. T. Delaney, S. P Danielsen, G. H. Fredrickson, and J. E. Shea, Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins, J. Phys. Chem. **10**, 1644 (2019).

[27] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, RNA length has a non-trivial effect in the stability of biomolecular condensates formed by RNA-binding proteins, PLoS Comput. Biol. **18**, e1009810 (2022).

[28] E. W. Martin and A. S. Holehouse, Intrinsically disordered protein regions and phase separation: Sequence determinants of assembly or lack thereof, Emerging Top. Life Sci. **4**, 307 (2020).

[29] M. P. Howard, Z. M. Sherman, A. N. Sreenivasan, S. A. Valenzuela, E. V. Anslyn, D. J. Milliron, and T. M. Truskett, Effects of linker flexibility on phase behavior and structure of linked colloidal gels, J. Chem. Phys. **154**, 074901 (2021).

[30] M. P. Howard, R. B. Jadrich, B. A. Linquist, F. Khabaz, R. T. Bonnecaze, D. J. Milliron, and T. M. Truskett, Structure and phase behavior of polymer-linked colloidal gels, J. Chem. Phys. **151**, 124901 (2019).

[31] M. P. Howard, Z. M. Sherman, D. J. Milliron, and T. M. Truskett, Wertheim's thermodynamic perturbation theory with double-bond association and its application to colloid–linker mixtures, J. Chem. Phys. **154**, 024905 (2021).

[32] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz, SAFT: Equation-of-state solution model for associating fluids, Fluid Phase Equilib. **52**, 31 (1989).

[33] M. S. Wertheim, Fluids with highly directional attractive forces. I. Statistical thermodynamics, J. Stat. Phys. **35**, 19 (1984).

[34] P. Ronceray, Y. Zhang, X. Liu, and N. S. Wingreen, Stoichiometry controls the dynamics of liquid condensates of associative proteins, Phys. Rev. Lett. **128**, 038102 (2022).

[35] Y. Zhang, B. Xu, B. G. Weiner, Y. Meir, and N. S. Wingreen, Decoding the physical principles of two-component biomolecular phase separation, eLife **10**, e62403 (2021).

[36] A. G. T. Pyo, Y. Zhang, and N. S. Wingreen, Surface tension and super-stoichiometric surface enrichment in two-component biomolecular condensates, iScience **25**, 103852 (2022).

[37] O. Kimchi, E. M. King, and M. P. Brenner, Uncovering the mechanism for aggregation in repeat expanded RNA reveals a reentrant transition, Nat. Commun. **14**, 332 (2023).

[38] T. C. Taylor, A. Backlund, K. Bjorhall, R. J. Spreitzer, and I. Andersson, First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*, J. Biol. Chem. **276**, 48159 (2001).

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXLife.1.023013 for details on our parametrization of EPYC1 from all-atom simulations; snapshots of phase separation in systems with stoichiometries of 3:1 and 10:1 EPYC1 to Rubisco stickers; the transition to a gas of rods at 1:1 EPYC1 to Rubisco sticker stoichiometry; a snapshot of simulations of Rubisco with movable stickers at 2:1 EPYC1 to Rubisco sticker stoichiometry; the probability of the presence of a sticker-sticker bond in the dense and dilute phases as a function of EPYC1 sticker location for the system shown in Fig. 3. The Supplemental Material also contains Refs. [23,40,41].

[40] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Saverdoti *et al.*, Scalable algorithms for molecular dynamics simulations on commodity clusters, in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (IEEE, New York, 2006), pp. 84–96.

[41] S. Piana, P. Robustelli, D. Tan, S. Chen, and D. E. Shaw, Development of a force field for the simulation of single-chain proteins and protein–protein complexes, J. Chem. Theory Comput. **16**, 2494 (2020).

[42] M. V. Turkina, A. Blanco-Rivero, J. P. Vainonen, A. V. Vener, and A. Villarejo, CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*, Proteomics **6**, 2693 (2006).

[43] C. S. Sørensen, A. Jendroszek, and M. Kjaergaard, Linker dependence of avidity in multivalent interactions between disordered proteins, J. Mol. Biol. **431**, 4784 (2019).

[44] E. T. Mack, P. W. Snyder, R. Perez-Castillejos, B. Bilgicer, D. T. Moustakas, M. J. Butte, and G. M Whitesides, Dependence of avidity on linker length for a bivalent ligand–bivalent receptor model system, J. Am. Chem. Soc. **134**, 333 (2012).

[45] M. Kjaergaard, J. Glavina, and L. B. Chemes, Predicting the effect of disordered linkers on effective concentrations and avidity with the "C_{eff} calculator" app, Methods Enzymol. **647**, 145 (2021).

[46] C. S. Sørensen and M. Kjaergaard, Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics, Proc. Natl. Acad. Sci. USA **116**, 23124 (2019).

[47] F. C. Mascali, R. Crespo, L. C. Tabares, and R. M Rasia, Conserved linker length in double dsRBD proteins from plants restricts interdomain motion, J. Magn. Reson. Open **16**, 100109 (2023).

[48] N. S. Gonzalez-Foutel. J. Glavina, W. M. Borcherds, M. Safranchik, S. Barrera-Vilarmau, A. Sagar, A. Estana, A. Barozet, N. A. Garrone, G. Fernandez-Ballester *et al.*, Conformational buffering underlies functional selection in intrinsically disordered protein regions, Nat. Struct. Mol. Biol. **29**, 781 (2022).

[49] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Chem. Phys. **117**, 1 (1995).

[50] K. Kremer and G. S. Grest, Dynamics of entangled linear polymer melts: A molecular-dynamics simulation, J. Chem. Phys. **92**, 5057 (1990).

[51] https://github.com/tgrandpr/Pyrenoid_cluster.git.

[52] R. J. Lunbeck and A. J. H. Boerboom, On the second virial coefficient of gas mixtures, Physica **17**, 76 (1951).

[53] H. Gould and J. Tobochnik, *Statistical and Thermal Physics: With Computer Applications* (Princeton University Press, Princeton, NJ, 2021).

[54] S. R. K. Ainavarapu, J. Brujić, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li, and J. M. Fernandez, Contour length and refolding rate of a small protein controlled by engineered disulfide bonds, Biophys. J. **92**, 225 (2007).