Supplementary Information for

Phase-separating pyrenoid proteins form complexes in the dilute phase

Guanhua He[1], Trevor GrandPre[2,3], Hugh Wilson[4], Yaojun Zhang[3,5,6], Martin C. Jonikas[1,7*], Ned S. Wingreen[1,4*], Quan Wang[4,8*]

[1] Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

[2] Department of Physics, Princeton University, Princeton, NJ 08544, USA

[3] Center for the Physics of Biological Function, Princeton University, Princeton, NJ, USA

[4] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

[5] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA

[6] Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA

[7] Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA

[8] Present address: Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

[*] Correspondence: mjonikas@princeton.edu, wingreen@princeton.edu, quan.wang@nih.gov

Supplementary Note 1

Although we managed to minimize EPYC1 sticking to the surface by using the PEI-PEG coating, we were not able to completely eliminate Rubisco sticking in the experiments described in Fig. 2. We thus expect the real concentration of Rubisco in solution to be lower than the nominal concentration. We added horizontal error bars in Fig. 2e to reflect this uncertainly in Rubisco concentrations. Specifically, we assume that Rubisco adsorption saturates at about 10 nM.

Supplementary Note 2

To estimate the K_D, we used the quadratic binding equation below to fit the data in Fig. 2d. This is because in our binding assay, Rubisco's concentration is not significantly higher than EPYC1's, so we cannot assume that only a small fraction of Rubisco bound to EPYC1[1]. The fraction of Rubisco bound EPYC1 can be expressed as

$$\frac{[AB]}{A_0} = \frac{(A_0+B_0+K_d) - \sqrt{(A_0+B_0+K_d)^2 - 4A_0B_0}}{2A_0}$$

In our case:

AB: Rubisco-EPYC1 complex

A_0: Rubisco concentration

B_0: EPYC1-GFP concentration

Then, diffusion rates of EPYC1-GFP measured by FCS should be weighted average of unbound and bound EPYC1-GFP, as D_{EPYC1} and D_{complex}
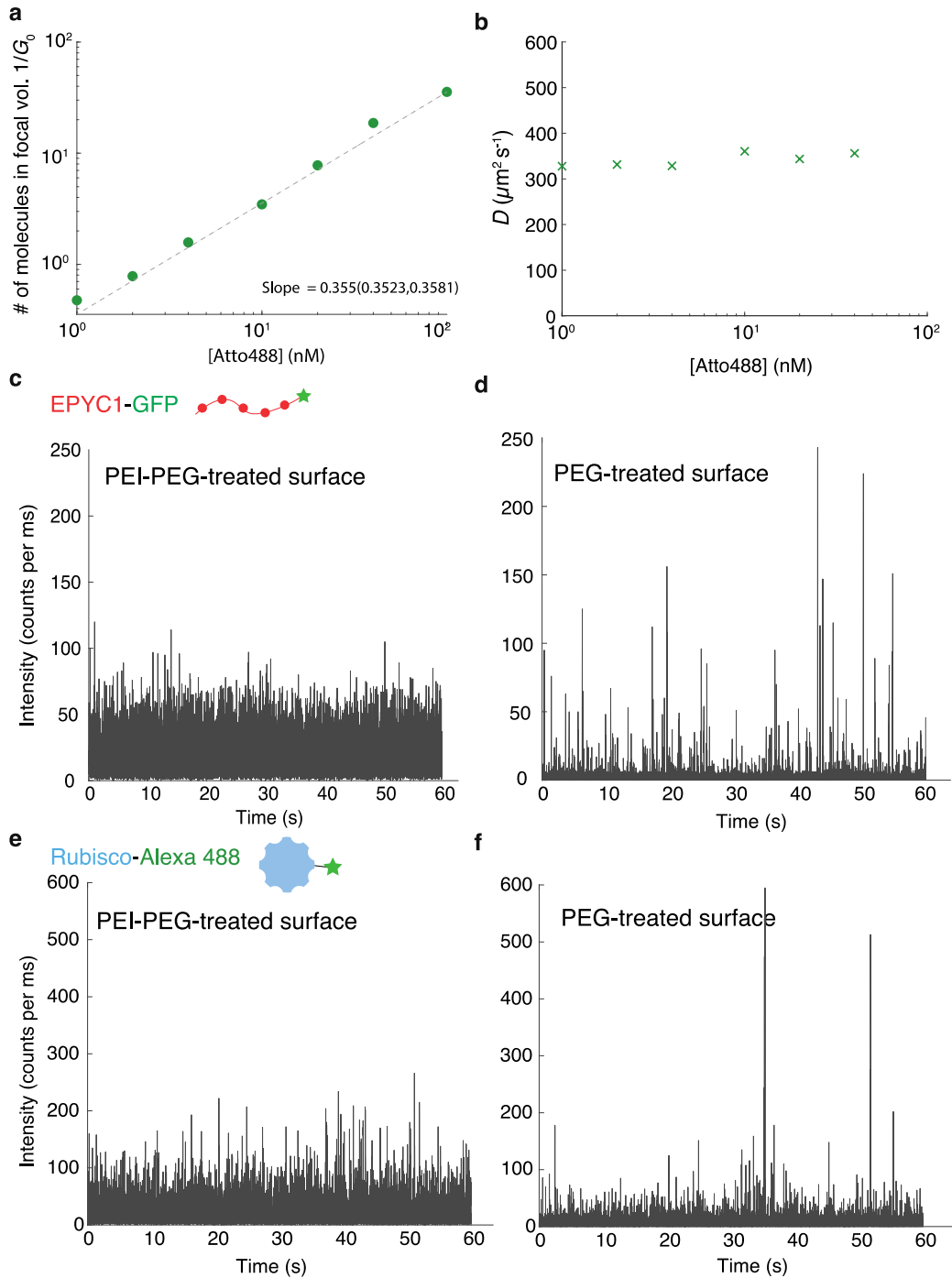
$$D = \frac{[AB]}{A_0} \cdot D_{\text{complex}} + \left(1 - \frac{[AB]}{A_0}\right) D_{\text{EPYC1}}$$

$$= D_{\text{complex}} \cdot \frac{(A_0+B_0+K_d) - \sqrt{(A_0+B_0+K_d)^2 - 4A_0B_0}}{2A_0}$$

$$+ D_{\text{EPYC1}} \cdot (1 - \frac{(A_0+B_0+K_d) - \sqrt{(A_0+B_0+K_d)^2 - 4A_0B_0}}{2A_0})$$

We used this equation to fit the data in Fig. 2d to extract K_d (unit: nM). We fixed $D_{EPYC1} = 62 \, \mu\text{m}^2/\text{s}$ and $D_{complex} = 41 \, \mu\text{m}^2/\text{s}$ to obtain the estimate of $K_d = 29 \pm 12$ nM (error bars represent 68% confidence interval):
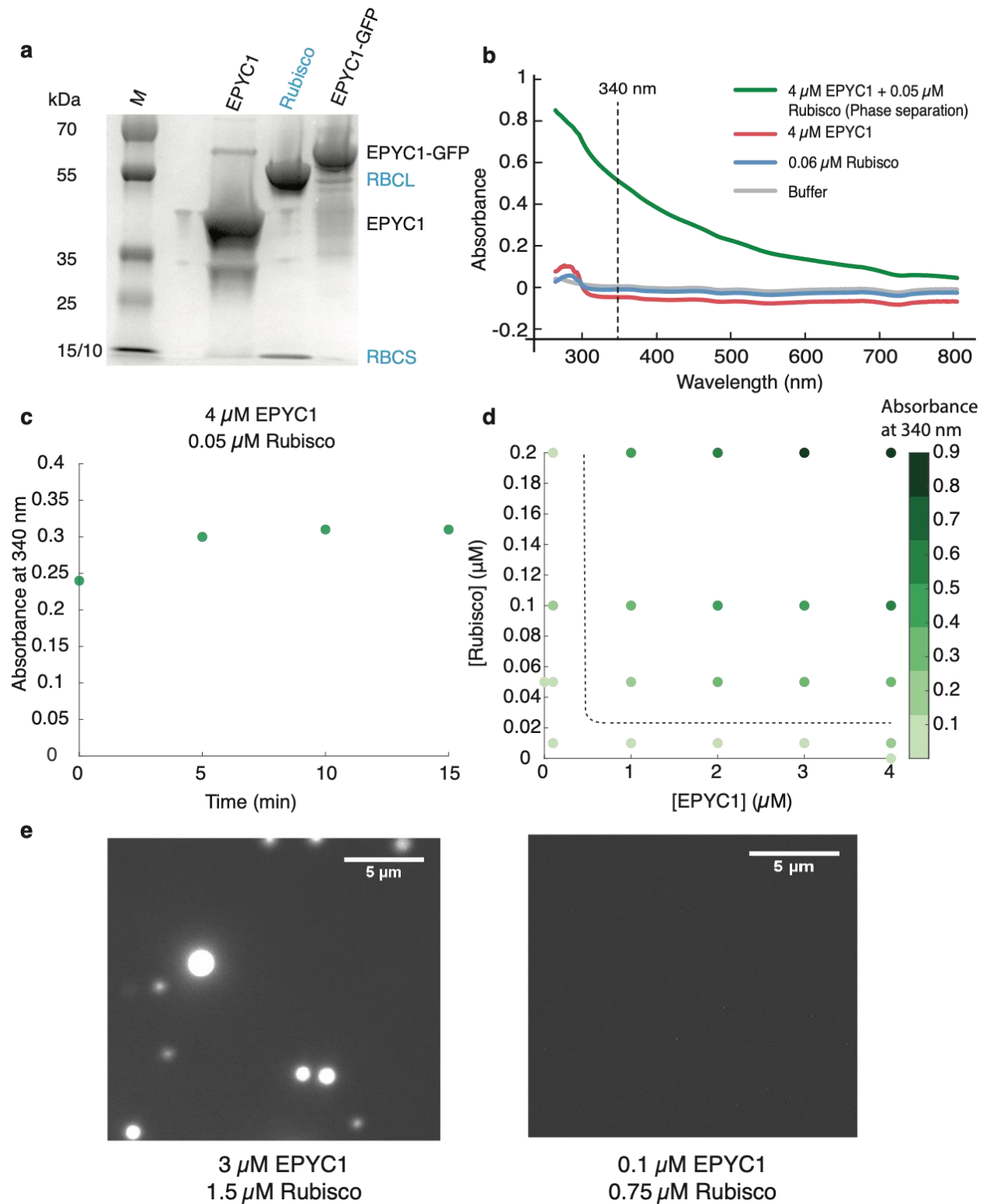
Supplementary Figure 1 | Use of a turbidity assay and fluorescence imaging to determine the phase diagram of EPYC1 and Rubisco. a Purified proteins on an SDS-PAGE gel. RBCL:

Rubisco large subunit. RBCS: Rubisco small subunit. M: marker. **b** UV/Vis extinction spectra of different sample solutions (the phase separated sample was measured 10 minutes after mixing). **c** The absorbance at 340 nm of a phase-separating solution consisting of 4 μM EPYC1 and 0.05 μM Rubisco as a function of time after mixing. **d** Absorption assay at 340 nm of mixed EPYC1-Rubisco solutions at concentrations shown on the x and y axes. **e** Representative fluorescence images of phase separation (left) and no phase separation (right). 20 nM EPYC1-GFP was added to both solutions.

a The number of Atto488 free dye molecules in the FCS focal volume were measured as a function of Atto488 concentrations. The focal volume was then calculated based on the slope of the fitted curve. **b** Atto488 diffusion rates were measured as a function of

Supplementary Figure 2 | FCS experiments were performed using a well-calibrated setup and an optimized surface. a The number of Atto488 free dye molecules in the FCS focal volume were measured as a function of Atto488 concentrations. The focal volume was then calculated based on the slope of the fitted curve. **b** Atto488 diffusion rates were measured as a function of

Atto488 concentrations. **c-d** FCS raw trace of 10 nM EPYC1-GFP on PEI-PEG-treated surface (**c**) and PEG-treated surface (**d**). **e-f** FCS raw trace of 10 nM Rubisco-Alexa 488 on PEI-PEG-treated surface (**e**) and PEG-treated surface (**f**). In panels **d** and **f**, the big spikes are protein aggregates, which were absent for PEI-PEG treated surfaces (panels **c** and **e**)
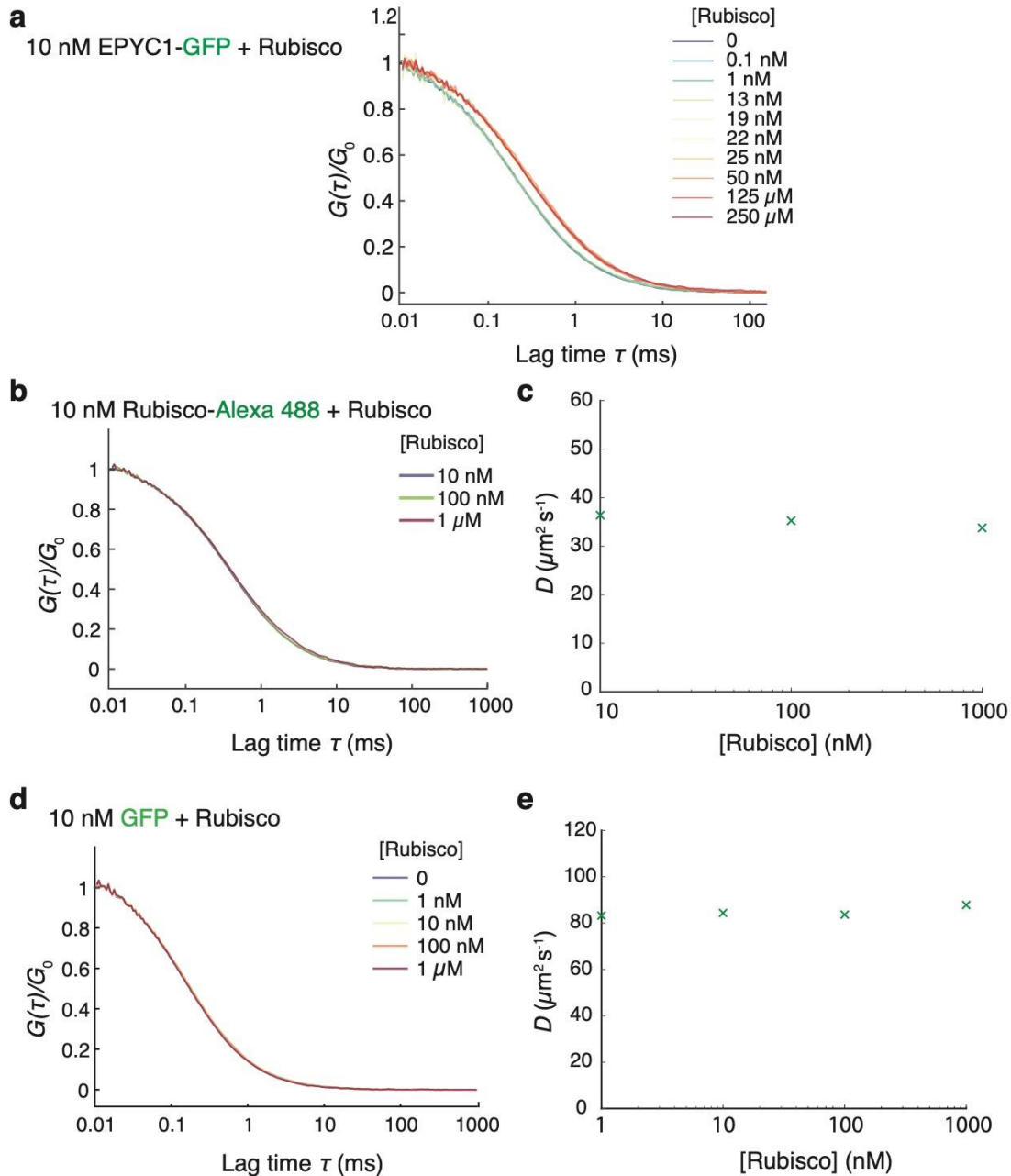
Supplementary Figure 3 | Neither interactions between GFP and Rubisco nor Rubisco self-interactions were observed in the FCS experiments. a Full autocorrelation curves of EPYC1-GFP fluorescence at different Rubisco concentrations. In **a**, **b**, **d**, each autocorrelation curve is normalized by its fitted value at $t = 0$ (Methods). **b** Autocorrelation curves of Rubisco labelled with Alexa Fluor 488 at different unlabeled Rubisco concentrations. **c** Diffusion rates of Rubisco-

Alexa 488 inferred from FCS data in **b** as a function of total Rubisco concentration. **d** Autocorrelation curves of GFP at different unlabeled Rubisco concentrations. **e** Diffusion coefficients of GFP inferred from FCS data in **d** as a function of Rubisco concentration.
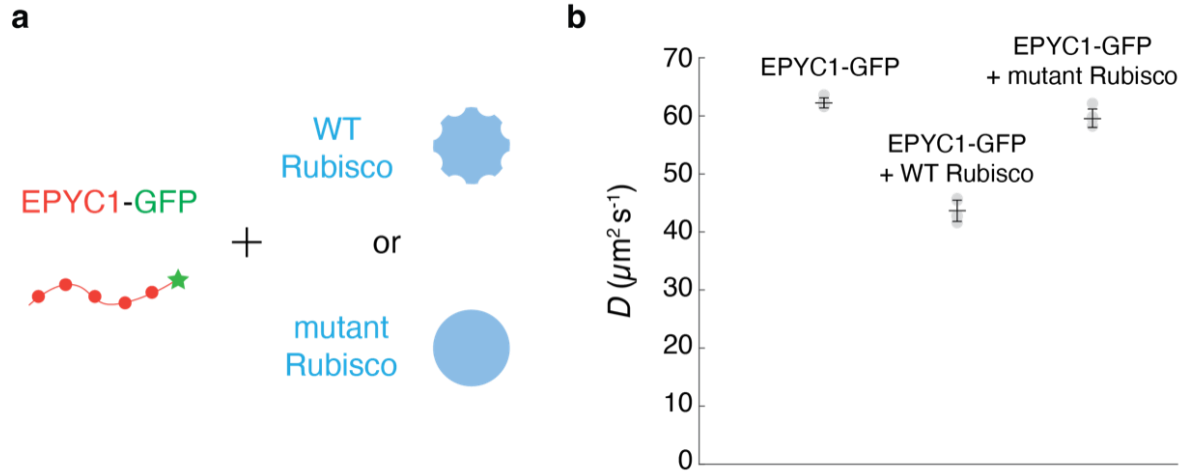
Supplementary Figure 4 | EPYC1 does not bind to a Rubisco mutant. a Cartoon showing the experiment setup: 10nM EPYC1-GFP is mixed with 13nM WT Rubisco or 13nM mutant Rubisco. The mutant Rubisco is M87D/V94D Rubisco small subunits from He et al. 2020 (Ref. 2). **b** Diffusion rates of EPYC1-GFP alone, with 13nM WT Rubisco, or 13nM mutant Rubisco. The error bars are standard deviations of repeated experiments (*n*=5).

a
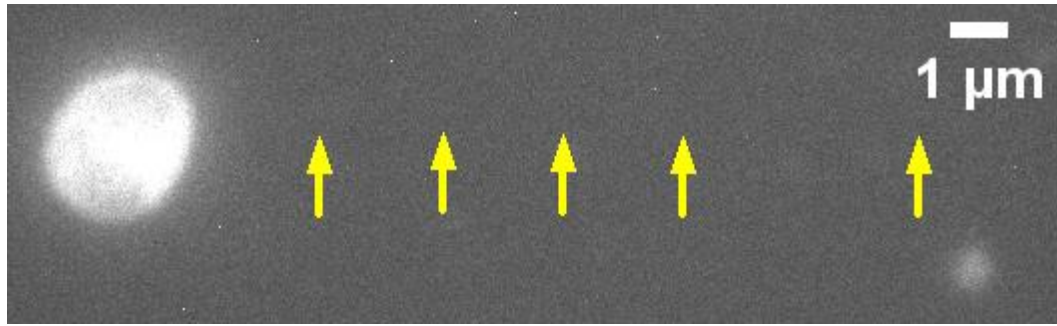
b
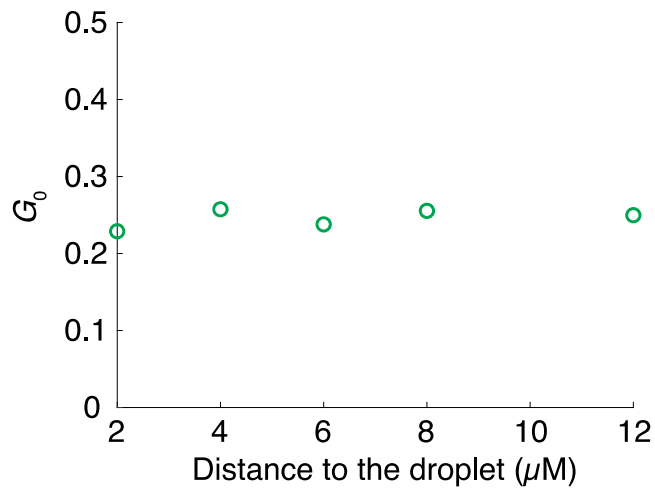
c
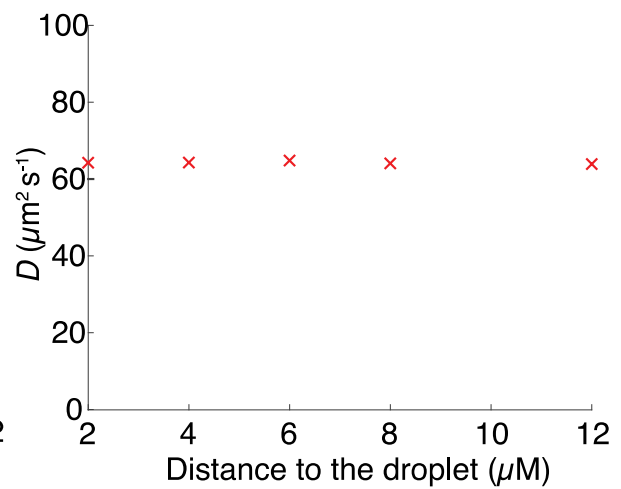
Supplementary Figure 5 | The measured dilute-phase properties are independent of measurement location. a FCS was performed in a focal volume 2, 4, 6, 8 and 12 μm away from the nearest droplet, as shown by yellow arrows. Bulk concentrations in the experiment: [EPYC1] = 5 μM, [Rubisco] = 0.25 μM, [EPYC1-GFP] = 40 nM. **b,c** Autocorrelation amplitude G_0 (**b**) and diffusion coefficient D (**c**) of EPYC1-GFP fitted from the FCS data is plotted against the distance to the droplet. Both extracted parameters are highly reproducible.

References:

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

2. He, S. *et al.* The Structural Basis of Rubisco Phase Separation in the Pyrenoid. *Nat. Plants* **6**, 1480–1490 (2020).