Electronic Supplementary Material (ESI) for Materials Advances.

Supporting Information for

Phase Transition of Recombinant Fusion Protein Assemblies and Vesicles in Macromolecularly Crowded Conditions

*Jooyong Shin, Yinhao Jia, Janani Sampath, Yeongseon Jang**

Department of Chemical Engineering, University of Florida, Gainesville, Florida 32611

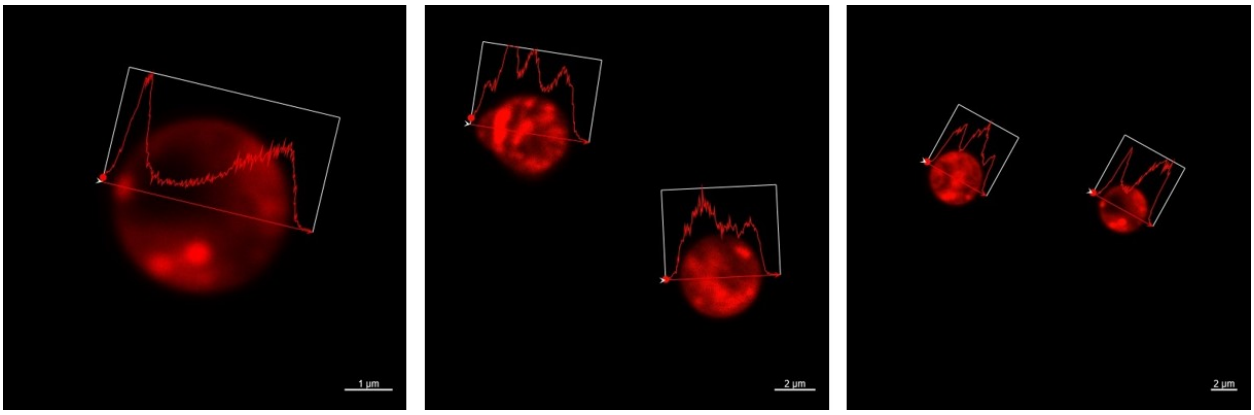

Figure S1. The intensity profiles of mCherry in the droplets assembled from the fusion proteins in PEG-rich solutions.

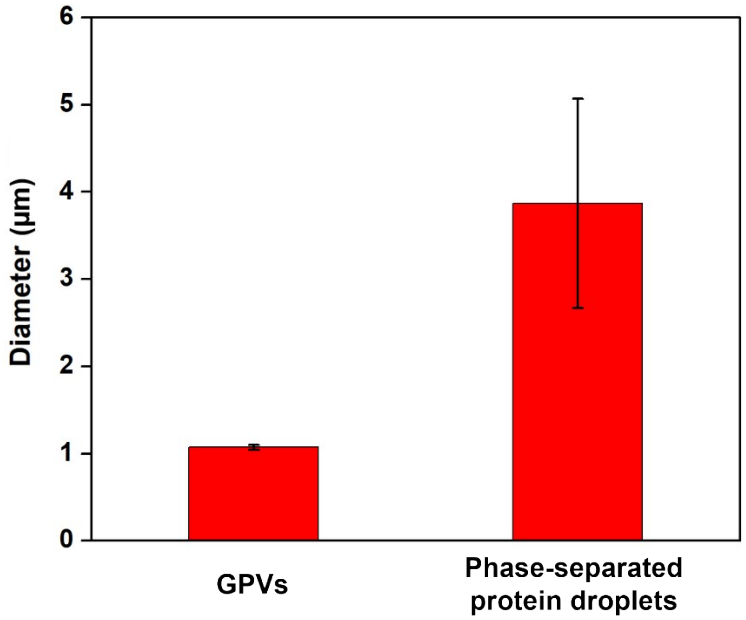

Figure S2. Comparative diameters of initially assembled mCherry-GPVs and crowding-induced phase-separated protein droplets. The statistical significance is denoted with **, indicating a p-value of less than 0.01

Simulation details

The molecule of mcherry and eGFP were taken from protein data bank, entry 2H5Q and 2WUR. The structure of ELP $((VPGVG)_2VPGFG(VPGVG)_2)_5$ was constructed based on the β-spiral configuration using angles adapted from previous literatures.[1] The chromophore group of the fluorescence globular proteins is replaced with the original amino acid sequence as the chromophore group is located inside the beta-barrel, and we do not investigate the mechanism of the fluorescence effect. Based on the preliminary simulations, which allow proteins to interact from different directions, the proteins prefer to interact and bind to each other according to the opposite charge of the surface. Thus, we are only interested in how the surface charge induces the free energy difference of the interactions between globular protein and ELP. CHARMM36m[2] force fields were used to represent the proteins along with the TIP3P water model.[3] Neutral termini capped 3 proteins ($-NH_2$ and -COOH) to avoid interactions introduced by terminal charge, as they are not the exact end of the proteins. All the systems were solvated in 12 x 12 x 12 nm^3 boxes with a total of ~170,000 atoms. Na^+ ions were introduced to neutralize each system. Hydrogen bonds were constrained during the simulation with the LINCS[4] algorithm. The cutoff for short-range electrostatic and Van der Waals interactions was 1.2 nm. The Particle Mesh Ewald (PME)[5] method was used to compute long-range electrostatic interactions; fast Fourier transform grid spacing was 0.15 nm. The VdW modifier was set to force-switch with VdW switch at 1.0nm. Constant temperature was maintained using the nose-hoover[6] algorithm. In constant-pressure simulations, the Parrinello-Rahman barostat[7] (1 bar, 2.0-ps coupling constant) was used. The time step for numerical integration was 2 fs.

The systems, prepared as described above, were first subject to steepest descent minimization with a maximum force of 1000 $kJ \cdot mol^{-1} \cdot nm^{-1}$, followed by two stages of 100 ps NVT equilibration. In the first stage, heavy atoms of the protein were restrained, and the solvent molecules were equilibrated. Initial velocities were randomly assigned according to Maxwell-Boltzmann distribution at 300K. In the second stage, all atoms were equilibrated with the velocity adopted from the first stage and position restraint is in place. Subsequently, a 1000 ps constant pressure and temperature simulation (NPT) simulation was conducted at 300K and 1bar with position restraint. The final outputs were used as the starting point for production simulation.

To prevent protein diffusion out of the simulation box and avoid the interaction between protein 2 and another side of protein 1, the UPPER_WALL option in PLUMED has been employed to limit the sampling to only the aligned direction of protein 1.

System	Tims(ns)	HH	BF	σ	CV	Pace(/ps)
eGFP-eGFP	300	1.5	10	0.13	d0	1
mCherry-mCherry	220	1.5	10	0.3	d0	1
ELP-ELP	200	1.5	10	0.35	d0	1
mCherry-ELP	200	1.5	10	0.38	d0	1

Table 1. simulation setup for well-tempered metadynamics. HH-Hill Height (kJ/mol), BF (Bias Factor), CV(Collective Variable), σ(kJ/mol). CV: Center-of-mass Distance

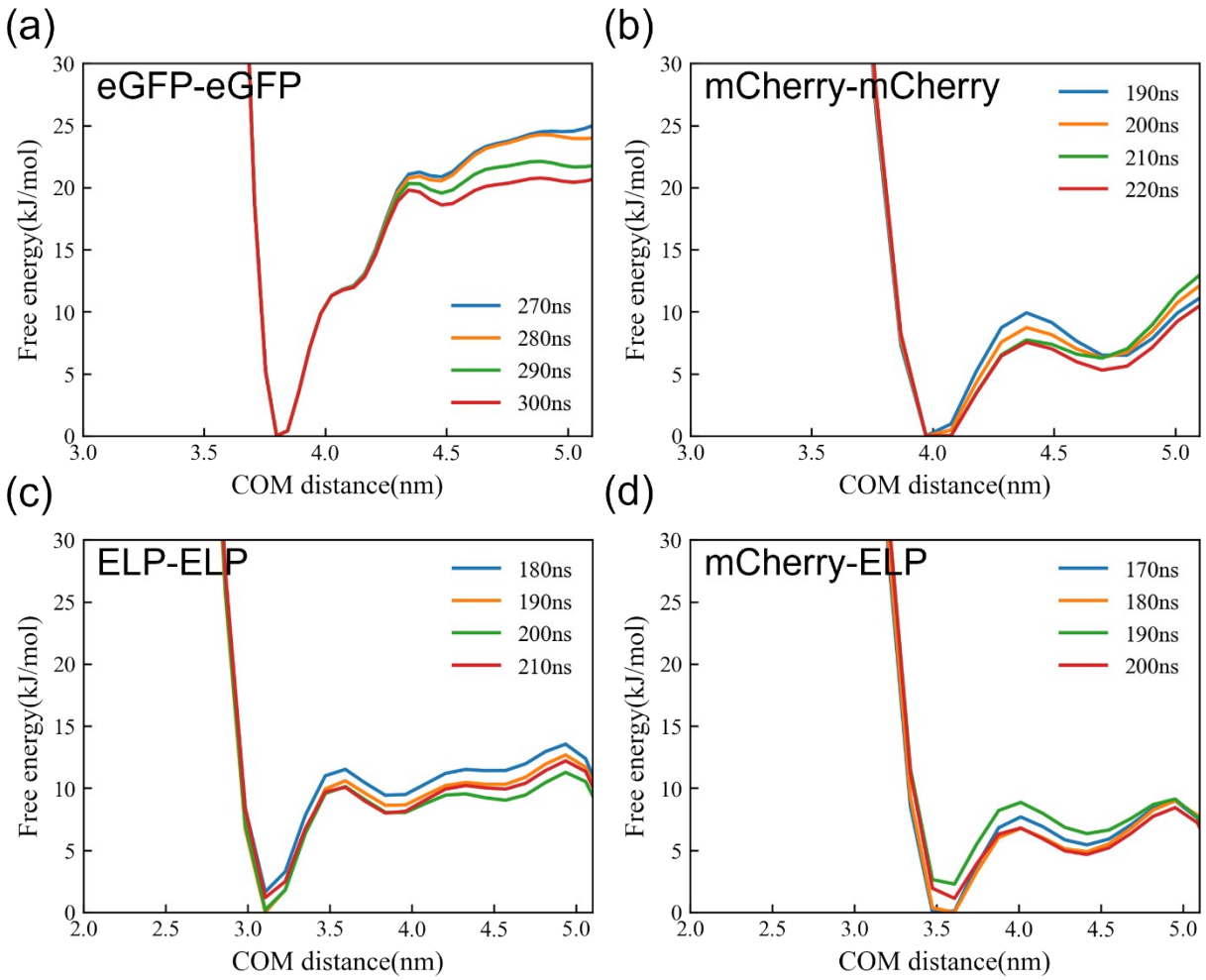

Figure S3. Convergence assessment showing last 40ns free energy profile change for (A) eGFP-eGFP, (B) mCherry-mCherry, (D) ELP-ELP, and (D) mCherry-ELP

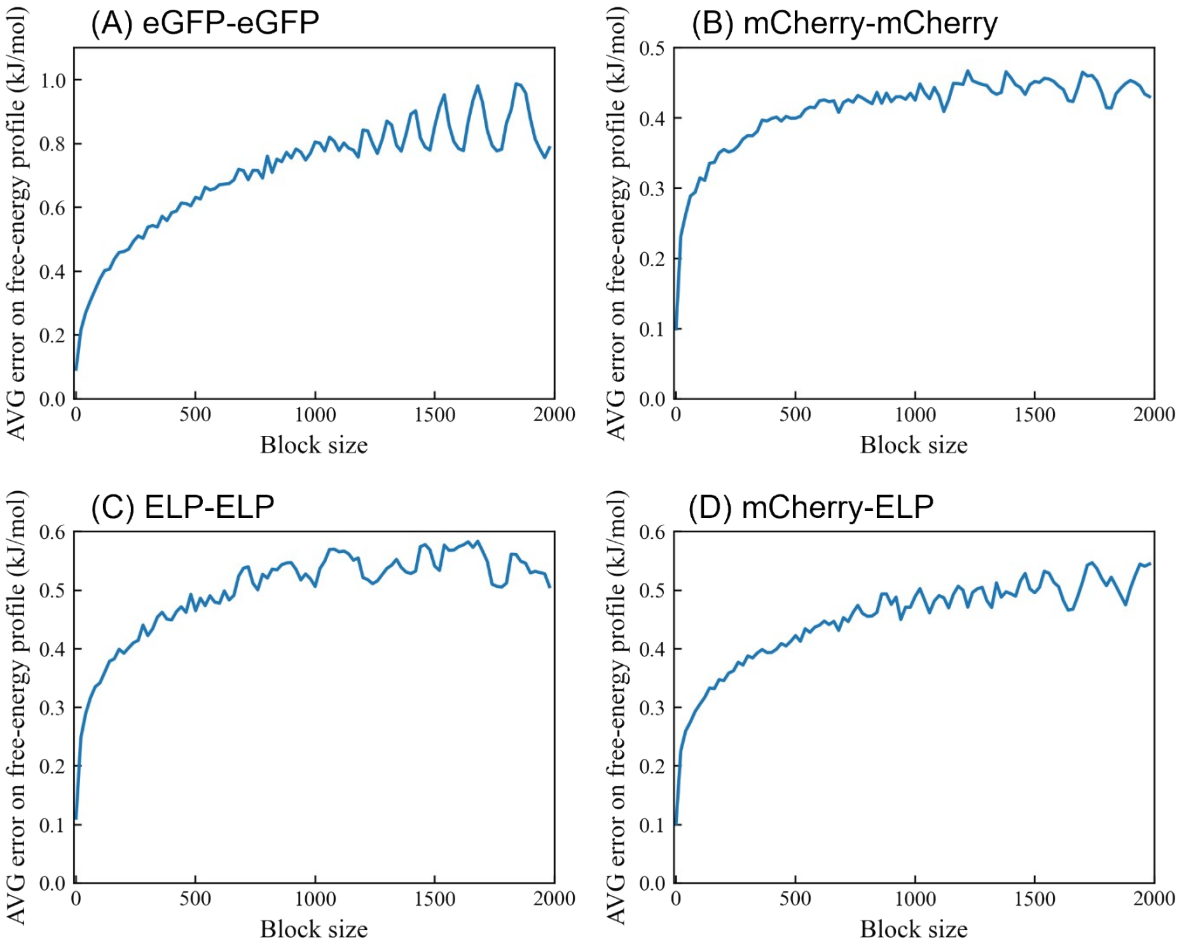

Figure S4. Convergence assessment showing average error on the free-energy profile with the varying block size for (A) eGFP-eGFP, (B) mCherry-mCherry, (C) ELP-ELP, and (D) mCherry-ELP.

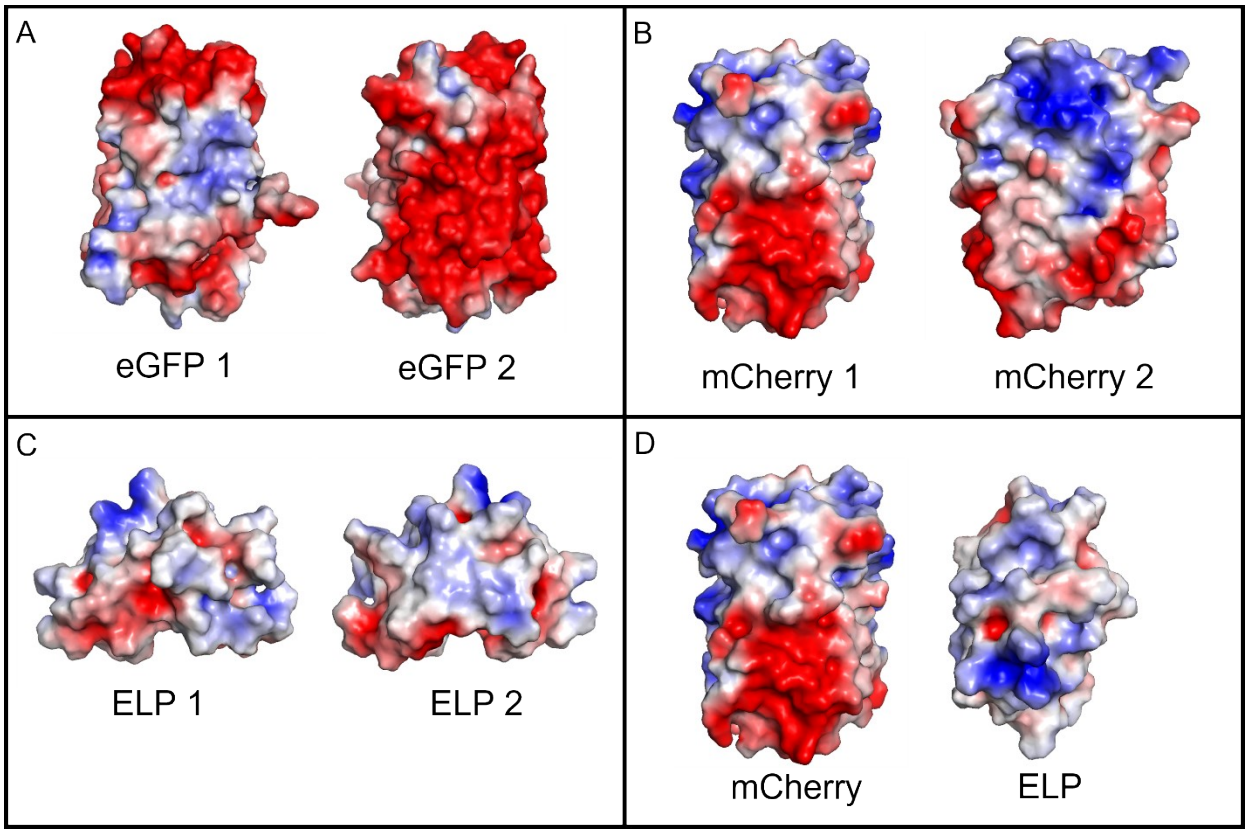

Figure S5. Simulation initial conformation alignment based on ABPS result with showing surface facing each other: (A) eGFP-eGFP, (B) mCherry-mCherry, (C) ELP-ELP, and (D) mcherry-ELP.

Reference

(1) Venkatachalam, C. M.; Urry, D. W. Development of a Linear Helical Conformation from Its Cyclic Correlate. β-Spiral Model of the Elastin Poly(Pentapeptide) (VPGVG)n. *Macromolecules* **1981**, *14* (5), 1225–1229. https://doi.org/10.1021/ma50006a017.

(2) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2016**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(4) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(5) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(6) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J Chem Phys* **1985**, *83* (8), 4069–4074. https://doi.org/10.1063/1.449071.

(7) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *Journal of Chemical Physics* **2007**, *126* (1). https://doi.org/10.1063/1.2408420.