

**Supporting Information for**  
**Conformational Properties of Polymers at Droplet Interfaces**  
**as Model Systems for Disordered Proteins**

Jiahui Wang,<sup>1</sup> Dinesh Sundaravadivelu Devarajan,<sup>1</sup> Arash Nikoubashman,<sup>2,\*</sup> and  
Jeetain Mittal<sup>1,3,4,\*</sup>

<sup>1</sup>Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station,  
TX 77843, United States

<sup>2</sup>Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz,  
Germany

<sup>3</sup>Department of Chemistry, Texas A&M University, College Station, TX 77843, United States

<sup>4</sup>Interdisciplinary Graduate Program in Genetics and Genomics, Texas A&M University, College  
Station, TX 77843, United States

\*Corresponding author email: [anikouba@uni-mainz.de](mailto:anikouba@uni-mainz.de), [jeetain@tamu.edu](mailto:jeetain@tamu.edu)

## Simulation Details

We have used a coarse-grained model, in which each monomer is represented by a spherical bead that is connected to neighboring beads via a harmonic spring:

$$U_b(r) = \frac{k_b}{2}(r - r_0)^2, \quad (1)$$

where  $r$  is the distance between two bonded beads,  $k_b = 20 \text{ kcal}/(\text{mol}\text{\AA}^2)$  is the spring constant and  $r_0 = 3.8 \text{ \AA}$  is the equilibrium length. The van der Waals interactions between nonbonded beads were modeling using a modified Lennard-Jones (LJ) potential<sup>1</sup>

$$U_{\text{vdW}}(r) = \begin{cases} U_{\text{LJ}}(r) + (1 - \lambda)\varepsilon, & r \leq 2^{1/6}\sigma \\ \lambda U_{\text{LJ}}(r), & \textit{otherwise} \end{cases} \quad (2)$$

where  $U_{\text{LJ}}$  is the standard LJ potential

$$U_{\text{LJ}}(r) = 4\varepsilon \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right], \quad (3)$$

with average hydrophathy  $\lambda = 0.865$  and bead diameter  $\sigma = 6.46 \text{ \AA}$  for the homopolymer systems.

The interaction strength was fixed to  $\varepsilon = 0.2 \text{ kcal/mol}$  in all simulations.

To have good statistics, droplet simulations were performed by placing 5000 chains (50 beads per chain) into a cubic box of edge length  $1010 \text{ \AA}$  with periodic boundary conditions in all directions. We performed Langevin dynamics simulations at temperature  $T = 300\text{K}$ . The friction coefficient was set as  $\gamma_i = m_i/t_{\text{damp}}$ , where  $m_i = 163.2 \text{ g/mol}$  and  $t_{\text{damp}} = 1000 \text{ fs}$ . Simulations were performed for  $2.5 \text{ }\mu\text{s}$  with a time step of  $10 \text{ fs}$  using HOOMD-blue<sup>2</sup> (ver. 2.9.3) with features

extended using azplugins<sup>3</sup> (ver. 0.10.2). The simulation trajectories were saved every 1 ns. Analysis was based on the last 1.5  $\mu$ s.

Bulk system simulations were performed by placing 500 chains into a cubic box at a constant pressure of  $P = 0$  atm for 0.5  $\mu$ s. After the chains achieved their preferred bulk-phase concentration, the simulation was performed using Langevin dynamics at constant volume for 1  $\mu$ s. All the simulations were implemented at a fixed temperature of  $T = 300$  K. The analysis was based on the last 0.5  $\mu$ s. Other simulation settings were the same with droplet simulation.

Single chain simulations were performed by placing one chain into a cubic box of edge length 160  $\text{\AA}$ . Langevin dynamics simulations were performed for 1  $\mu$ s at constant temperature  $T = 300$  K. The analysis was based on the last 0.5  $\mu$ s. Other simulation settings were the same with droplet simulation.

Natural protein simulations were performed by placing 1500 chains into a cubic box of edge length 1000  $\text{\AA}$  with periodic boundary conditions in all directions. Langevin dynamics simulations were performed at temperature  $T = 300$ K for FUS LC and  $T = 260$ K for LAF-1 RGG. The hydrophobicity values ( $\lambda_{ij}$ ) were set according to the HPS-Urry model<sup>4</sup>. The friction coefficient was set as  $\gamma_i = m_i/t_{\text{damp}}$ , where  $m_i$  was the residue mass and  $t_{\text{damp}} = 1000$  fs. Simulations were performed for 1  $\mu$ s and other simulation settings were the same with droplet forming simulation.

### **Natural protein sequences used in this work**

FUS LC (length = 163)

MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSSYSGYSQSTDTSGYGQSSYSSY  
 GQSQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGGQSSYPGYGQQPAPSSTSGSYGSSS  
 QSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNS

LAF-1 RGG (length = 168)

MESNQSNNGGSGNAALNRGGRYVPPHLRGGDGGAAAAASAGGDDRRGGAGGGGYRR  
 GGGNSGGGGGGGYDRGYNDNRDDRNRGGSGGYGRDRNYEDRGYNGGGGGGGNRG  
 YNNNRGGGGGGYNRQDRGDGSSNFSRGGYNNRDEGSDNRGSGRSYNNDRRDNGGD  
 G

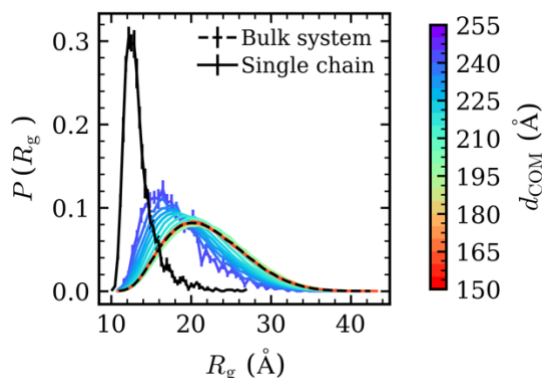
### Average angle calculation for isotropic distribution of segments

The average angle of segments distributed without any preferred orientation can be calculated as

$$\theta_{\text{ave}} = \int_{\pi/2}^{\pi} \theta \sin\theta \, d\theta = \pi - 1, \quad (4)$$

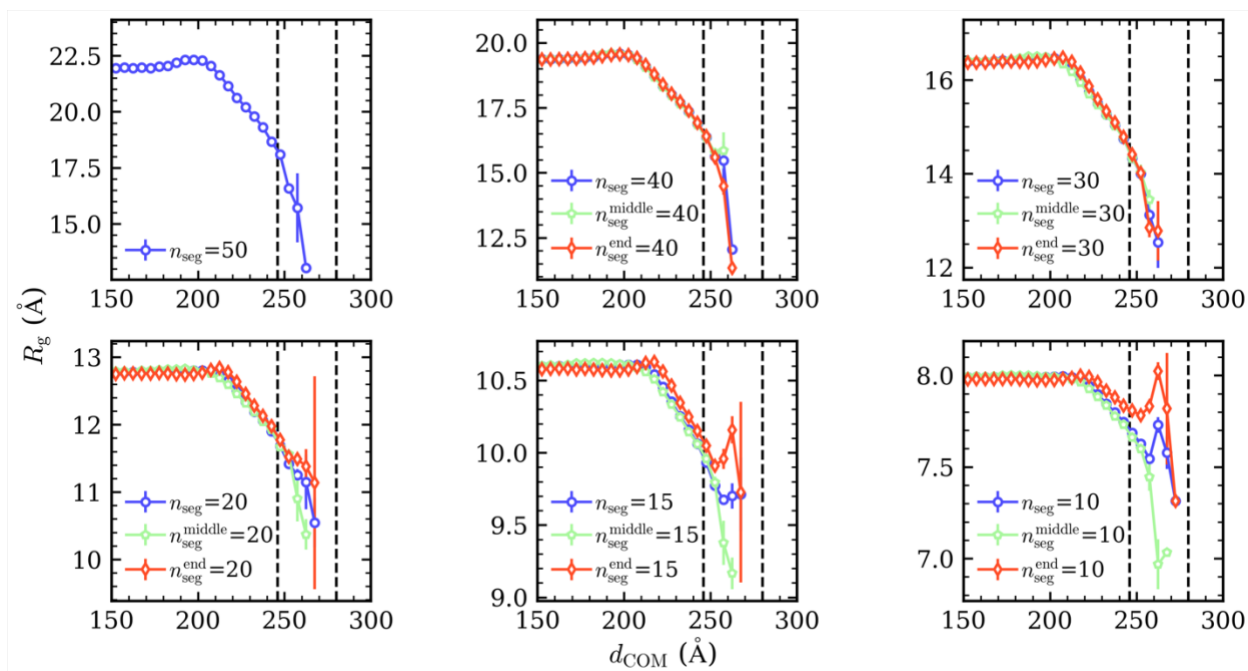
where  $\theta$  is the angle between the segment-to-droplet COM vector and the segment end-to-end vector.

### Supplementary Figures

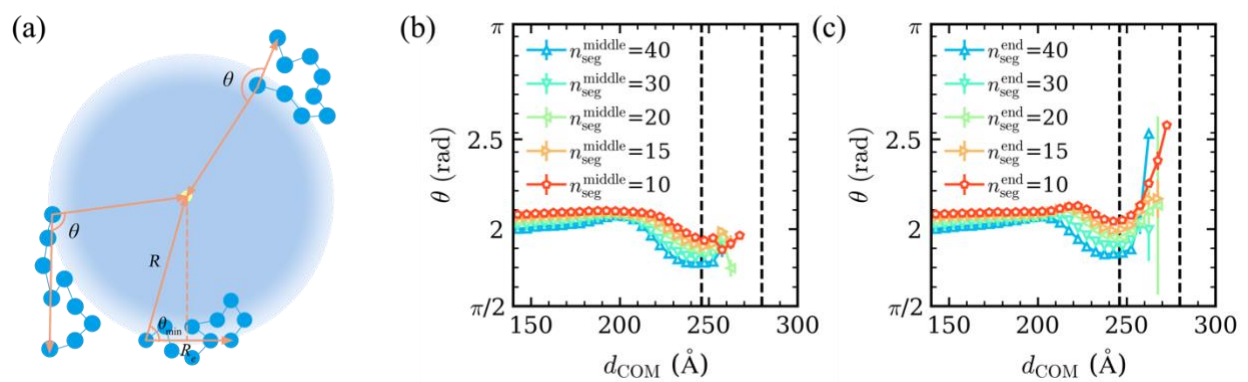


**Fig. S1.** Distribution of radius of gyration ( $R_g$ ). The black dashed line represents the  $R_g$  distribution in the bulk system. The black solid line represents the  $R_g$  distribution of a single chain. The color gradient

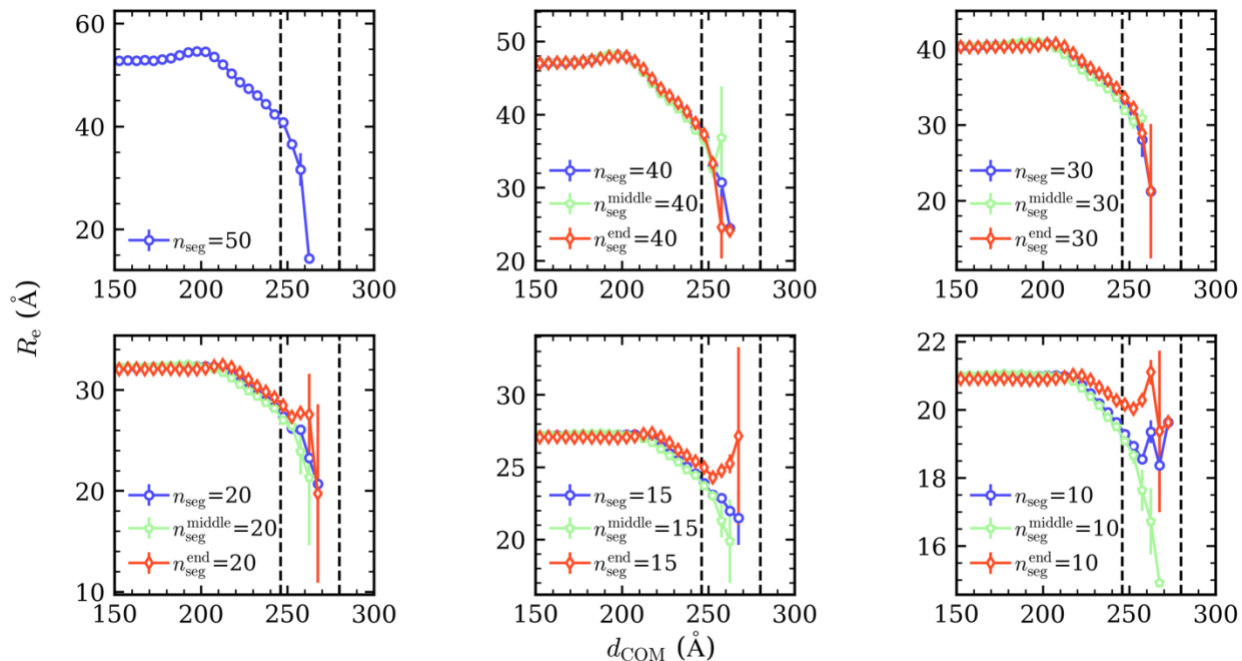
from red to purple corresponds to the distribution of  $R_g$  for chains with  $d_{\text{COM}}$  ranging from 150 to 255 Å.



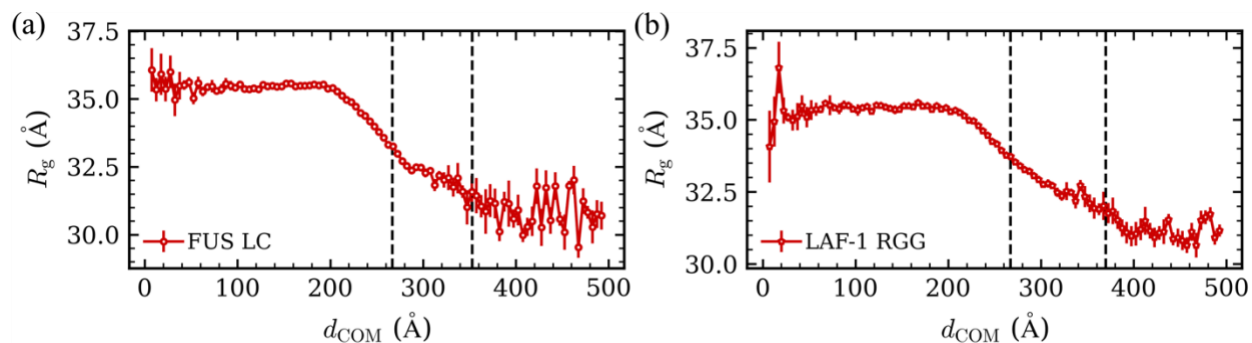
**Fig. S2.**  $R_g$  of chains and segments consisting of 40, 30, 20, 15, and 10 monomers as a function of  $d_{\text{COM}}$ . The purple line represents the average  $R_g$  for all segments. The green line represents segments located in the middle of the chain, while the red line represents segments located at the chain's ends. The black dashed lines represent the boundaries of the droplet interface.



**Fig. S3.** Orientation calculation. (a) Schematic of calculation method for angle ( $\theta$ ) and the minimum of  $\theta$  ( $\theta_{\text{min}}$ ). (b)  $\theta$  of different length segments located at the middle of a chain with respect to  $d_{\text{COM}}$ . (c)  $\theta$  of different length segments located at the ends of a chain with respect to  $d_{\text{COM}}$ .



**Fig. S4.** End-to-end distance ( $R_e$ ) of chains and segments consisting of 40, 30, 20, 15, and 10 monomers as a function of  $d_{\text{COM}}$ . The purple line represents the average  $R_e$  for all segments. The green line represents segments located in the middle of the chains, and the red line represents segments located at the chain's ends.



**Fig. S5.**  $R_g$  as a function of  $d_{\text{COM}}$  for (a) FUS LC and (b) LAF-1 RGG sequences. The black dashed lines represent the boundaries of the interface.

## REFERENCES

- (1) Ashbaugh, H. S.; Hatch, H. W. Natively Unfolded Protein Stability as a Coil-to-Globule Transition in Charge/Hydrophathy Space. *J. Am. Chem. Soc.* **2008**, *130* (29), 9536–9542.
- (2) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-Blue: A Python Package for High-Performance Molecular Dynamics and Hard Particle Monte Carlo Simulations. *Comput. Mater. Sci.* **2020**, *173*, 109363.
- (3) See <https://github.com/mphowardlab/azplugins> for source code for the software.
- (4) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved Coarse-Grained Model for Studying Sequence Dependent Phase Separation of Disordered Proteins. *Protein Sci.* **2021**, *30* (7), 1371–1379.