Supplementary Information for:

Single fluorogen imaging reveals spatial inhomogeneities within biomolecular condensates

Authors: Tingting Wu[1,2], Matthew R. King[2,3], Mina Farag[2,3], Rohit V. Pappu[2,3*], Matthew D. Lew[1,2*]

Affiliations:

[1] Department of Electrical and Systems Engineering, Washington University in St. Louis, James F. McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

[2] Center for Biomolecular Condensates, James F. McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

[3] Department of Biomedical Engineering, Washington University in St. Louis, James F. McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

e-mail: pappu@wustl.edu, mdlew@wustl.edu

Table of Contents

Supplementary Methods

Sample Preparation

Expression and purification of A1-LCD. BL21-CODON Pluse RIPL *E. coli* cells (Agilent Technologies 230280) transfected with Δhexa_His-TEV-A1-LCD [1] were grown in LB broth (Sigma: L3522) in Erlenmeyer flasks with ≥5-fold head volume at 37˚C and 220 rotation per minute (RPM) orbital shaking until OD_{600} ~0.6 was reached. Cultures were then chilled for 15 minutes in an ice bath before being in induced with 0.5mM IPTG. Expression-induced cultures were grown for 18 hours at 18˚C. Cells were harvested by centrifugation, washed of residual media, and stored as pellets in 50mL Falcon tubes at -80˚C.

At the start of a purification, the cell pellet was gently resuspended to homogeneity in 35 mL supplemented lysis buffer (50 mM MES, 500 mM NaCl, 14.3 mM BME (β-Mercaptoethanol), 200 µM PMSF (phenylmethylsulfonyl fluoride) pH 6), then lysed via sonication on a Branson 550 with an L102C horn attachment using five series of the following 20-round cycle: 1 second on / 2 second off at 30% power. Lysis buffer supplements are: 500U DNAaseI (Sigma - 4536282001), 500U RNAse A (Sigma - 10109169001), 5 mg Lysozyme (Sigma - 62971-10G-F), and one protease inhibitor tablet (Sigma - 40694200). After a 30-minute spin at 38,000g, the pellet was resuspended in a resuspension buffer (6 M GdmCl, 20 mM Tris, 15 mM imidazole, 14.3mM BME, and 200 µM PMSF; pH 7.5) via repeat pipetting and pulses of sonication until the solution was visibly homogenous. This solubilized pellet was spun at 38,000g for 30 minutes and the supernatant was applied to a gravity column with 5 mL bed volume of HiPur NiNTA resin (Fisher – 88223) equilibrated with resuspension buffer. The NiNTA resin was washed in 75mL wash Buffer (20 mM Tris, 30mM ImZ, 4 M Urea, 14.3mM BME, and 200 µM PMSF; pH 7.5), then protein eluted in elution buffer (20 mM Tris, 350mM ImZ, 4 M Urea, 14.3mM BME, and 200 µM PMSF; pH 7.5). Peak fractions from NiNTA affinity purification were pooled and diluted 1:1 in dilution buffer (20 mM Tris and 14.3 mM BME; pH 7.5). To this, 1 mg of TEV protease was added and the mixture was dialyzed overnight in dialysis buffer (20 mM Tris, 2 M urea, 50 mM NaCl, 0.5 mM EDTA, and 1 mM DTT; pH 7.5). The following morning, the protein solution was filtered through a 0.22 µm filter, then further purified via ion exchange chromatography on an ÄKTA Pure fast protein liquid chromatography (FPLC) module using a HiTrap SP 5mL column (Cytiva – 17115201). The column was equilibrated with buffer A (20 mM Tris, 2 M urea, 50 mM NaCl, and 14.3 mM BME; pH 7.5) and Protien was bound then subjected to a contiguous gradient protocol with from 0.05M NaCl to 1 M NaCl. Fractions containing A1-LCD were pooled purified to ≥99% purity using size exclusion chromatography - HiLoad 16/600 Superdex 200pg column (Cytiva - 28989335) on the ÄKTA Pure FPLC module in storage buffer (20mM MES and 4 M GdmCl at pH 5.5). A1-LCD fractions were pooled and concentrated in Amicon Ultra 3 MWCO (molecular weight cut-off) concentrator columns (Millipore-Sigma UFC 500396), according to the manufacturer's suggestions. Concentrated protein was stored at 4˚C until used.

After each stage of purification (affinity, ion exchange, and/or size exclusion), protein concentration and nucleotide contamination were assessed on a Nanodrop2000 via absorbance at 280 nm measurements and A260/A280 measurements. Similarly, after each purification stage the inputs, flow-throughs, washes, and elutes were interrogated on using SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) (Gel: BioRad miniPROTEAN TGX AnyKD - 4569036; MW latter: BioRad Precession Plus Protein Standard Unstained – 1610363), stained with

EZblue Coomassie stain (10% (v/v) phosphoric acid, 10% (w/v) Ammonium sulfate, 20% (v/v) Methanol, 1.2% (w/v) Coomassie Blue) and de-stained via serial washes in ddH_2O.

Expression and purification of DDX4-NT. BL21 *E. coli* cells (NEB - C2530H) transfected with His-SUMO-DDX4-NT [2], a gift from Professor Lewis E. Kay, were grown in the same manner as A1-LCD. Cells were lysed in the same manner as A1-LCD, with the only difference being the supplemented lysis buffer (20 mM Sodium Phosphate, 750 mM NaCl, 20 mM Imidazole, 14.3 mM BME, and 0.2 mM PMSF; pH 7.5). Supernatant was recovered from a 25-minute spin at 38,000g and bound to an equilibrated HisTrap FF Crude 5mL column (Cytiva – 11000458) using an ÄKTA Pure fast protein liquid chromatography (FPLC) module. The NiNTA column was washed in 75 mL lysis buffer then protein was eluted in elution buffer (0.02M Sodium Phosphate, 0.5 M NaCl, 0.35 M Imidazole, 0.0143 M BME, 0.0002 M PMSF; pH 7.5). Peak fractions from this affinity purification were pooled and diluted 5-fold in dilution buffer (0.02 M Sodium Phosphate, 0.0143 M BME; pH 7.5). This solution was further purified via ion exchange chromatography using a contiguous gradient purification protocol with a HiTrap Heparin HP 5mL column (Cytiva – 17040703), Buffer A (0.02M Sodium Phosphate, 0.1M NaCl, 0.0143M BME; pH 7.5), and Buffer B (0.02 M Sodium Phosphate, 0.1 M NaCl, 0.0143 M BME; pH 7.5) on the ÄKTA Pure FPLC module. Peak fractions containing were pooled and cleaved of SUMO tags during an overnight dialysis in the presence of 0.02x ULP1 Protease in cleavage buffer (0.02 M Sodium Phosphate, 0.3 M NaCl, 0.001M DTT (Dithiothreitol), pH 7.5). DDX4-NT was purified to ≥98% using size exclusion chromatography - HiLoad 16/600 Superdex 200pg column (Cytiva - 28989335) on the ÄKTA Pure FPLC module in storage buffer (0.022M Sodium Phosphate, 1.1M NaCl, 0.0143M BME, pH 7.5). The DDX4-NT solution was supplemented with 10% glycerol and concentrated in Amicon Ultra 3 MWCO (molecular weight cut-off) concentrator columns (Millipore-Sigma UFC 500396) and concentrated protein was aliquoted into single-use volumes (typically 10 µL), flash froze in liquid N_2, and stored at -80˚C. Each step of the purification was assessed in the same manner as described for A1-LCD.

Preparation of condensates for imaging. We adhere a silicone isolator (Grace Bio-labs, SKU 665206) to a coverslip (DeckGlaser coverglass, 24×60 mm, 170±5 µm) to create a small chamber for imaging. To induce phase separation, we mixed 9 µL of ~ 100 µM A1-LCD (in 20 mM HEPES buffer, pH 7) with 1 µL of salt buffer (20 mM HEPES buffer with 3 M NaCl, pH 7.0) into the chamber to a final concentration of ~ 90 µM A1-LCD in 20 mM HEPES buffer with 300 mM NaCl. We then immediately added 0.5 µL of a fluorogenic probe solution (20 mM HEPES buffer with 300 mM NaCl, pH 7.0) to the chamber. The final concentrations of each probe are shown in **Supplementary Table 2**. Phase separation occurs spontaneously within the chamber at room temperature (22° C) to form condensates for imaging. We noticed that condensates of His-tagged A1-LCD were too hydrophobic to completely wet the coverslip. We therefore use A1-LCD for the wetting experiment shown in **Figure 1e-g** and **Figure 2**, while an A1-LCD mixture containing ~20% His-tagged A1-LCD was used for experiments requiring well-structured condensates shown in **Figure 1b-d, Figure 3, and Figure 4**. Notably, the features of these various systems as measured by NB, NR, and MC540 are identical (e.g., **Fig. 1b-d** compared to **Fig. 1e-g)**. To form DDX4 condensates for imaging, we mixed 1 µL of ~ 58 mg/µL DDX4 (in 8 µM NaH_2PO_4, 12 µM Na_2HPO_4, 0.75 M NaCl) with 9 µL low-salt solution (8 µM NaH_2PO_4, 12 µM Na_2HPO, 0 M NaCl) into a small chamber for imaging.

Optical microscope setup and imaging procedure

Optical setup. We used a home-built inverted microscope for epifluorescence imaging, single-molecule imaging, and single-molecule tracking (Error! Reference source not found.). The microscope is equipped with an oil-immersive objective (OLYMPUS UPLSAPO100XOPSF, NA 1.4). The excitation laser, dichroic mirror, and emission filters were switched to match the excitation and emission spectra of NB, NR, and MC540 (**Supplementary Table 3**). A relatively low peak laser intensity (~ 51 W/cm^2) and high concentration of fluorogenic probes were used for epifluorescence imaging (**Supplementary Table 2**). A relatively large peak laser intensity ~ 4088 W/cm^2 coupled with a low concentration of fluorogenic probes facilitates bright and bright fluorescence flashes for single-molecule imaging and tracking.

Imaging protocols using multiple fluorogenic probes. As the emission and excitation spectra of MC540 and NB are well separated, we imaged single condensates using multiple probes by switching the excitation laser, dichroic mirror, and emission filter appropriately without crosstalk (**Supplementary Table 3**). However, there is crosstalk between the emission spectra of NB and NR. The emission spectrum of NR significantly overlaps with the red window of our microscope (676$\pm$18 nm, **Fig. 2d**), while the emission spectrum of NB is barely observable in the orange window (593$\pm$23 nm, **Fig. 2c**). Thus, to image a single condensate using both NB and NR, we always captured NB images first in the absence of NR using the red-pass filter, and then added NR to image the condensates using the orange-pass filter (**Supplementary Table 3**).

Data analysis

Single-molecule detection and estimation algorithm. Single-molecule imaging and tracking requires detecting fluorescence "flashes" within raw camera images and estimating their 2D positions. We use a bespoke regularized maximum likelihood estimator, RoSEClick or tap here to enter text. [3] to estimate SM positions, brightness, and fluorescence burst durations. A regularization parameter of 0.3 and a calibrated point-spread function [4] model is used for estimation to account for optical aberrations.

Single-molecule tracking analysis – grouping localizations and quantifying burst durations and displacement trajectories. At a low concentration of fluorogenic probes, emitters are sparse enough to be easily distinguishable from one another. Two localizations are classified as originating from a single emitter if they appear within two adjacent camera frames and are separated by a distance smaller than a threshold T = $175 + 4\sigma_r$ nm, where σ_r is the localization precision that varies for emitters with different signal-to-background ratios (~ 15 nm on average). Localizations collected at 10 ms intervals (the exposure time of the camera) are grouped together into a single trajectory until an emitter becomes dark, namely, when all emitters in the next frame lie at a distance > T from the current emitter. The burst duration of a fluorescence emitter is equal to $N_f \times 10$ ms, where N_f is the number of frames in the trajectory (**Fig. 4b**). The magnitude of the velocity of an emitter is calculated as the Euclidean distance between its positions in each pair of consecutive frames within a trajectory (**Fig. 4e,h,i**); if the trajectory is longer than two frames, each calculated speed is reported separately (e.g., in **Fig. 4e**).

Quantifying the degree of clustering within each condensate. The cluster coefficient is calculated for each localization based on Ripley's $H(r)$ and Getis and Franklin's $L(r)$ functions [5,6]. To assign a cluster coefficient to the i^{th} localization p_i, we define a circular region centered at p_i with a radius r of 50 nm. The number of single molecules within this circular region is used to quantify:

$$H(r) = \sqrt{\frac{N_{p_i}}{\lambda\pi}} - r;$$

Here, N_{p_i} represents the number of SMs within the circular region and λ is the average localization density across each condensate. Localizations are classified as clustered if $H(r)$ is above some threshold (e.g., 20 in **Fig. 3d**).

Excess variance. We note that if single molecule localizations are described by a homogenous Poisson process, otherwise referred to as complete spatial randomness, then we expect the mean localization density to match the variance of the localization density when both statistics are calculated across an entire condensate (**Extended Data Fig. 4**). Thus, the heterogeneity of localizations within a condensate may be quantified using an excess variance, defined as:

$$V = \frac{\text{variance}(S_{con})}{\text{mean}(S_{con})} - 1;$$

Here, S_{con} represents a collection of localization densities (per 20 nm × 20 nm bin) measured within a condensate.

Coordinate-based correlation (CBC) between fluorogenic probes. We quantified the colocalization between pairs of fluorogenic probes (A and B) directly using the position coordinates provided by SMLM rather than the brightness of pixels within an image [7]. To calculate the CBC value for a specific localization A_i from dataset A, the distribution of localizations from both species A and B around A_i is calculated as:

$$D_{A_i,A}(r) = \frac{N_{A_i,A}(r)}{N_{A_i,A}(R_{max})}\frac{R_{max}^2}{r^2}$$

$$D_{A_i,B}(r) = \frac{N_{A_i,B}(r)}{N_{A_i,B}(R_{max})}\frac{R_{max}^2}{r^2}$$

where $N_{A_i,A}(r)$ and $N_{A_i,B}(r)$ are the number of localizations of species A and species B within the distance r around A_i respectively. In our calculation, we set the R_{max} to be 300 nm, which determines the size of the region over which correlations are calculated. We then calculate the rank correlation coefficient as

$$S_{A_i} = \frac{\sum_{\mathrm{r_j}=0}^{\mathrm{R_{max}}}\left(\mathrm{O_{D_{A_i,A}}(r_j)} - \overline{\mathrm{O}}_{\mathrm{D_{A_i,A}}}\right)\left(\mathrm{O_{D_{A_i,B}}(r_j)} - \overline{\mathrm{O}}_{\mathrm{D_{A_i,B}}}\right)}{\sqrt{\left(\sum_{r_j=0}^{\mathrm{R_{max}}}\left(\mathrm{O_{D_{A_i,A}}(r_j)} - \overline{\mathrm{O}}_{\mathrm{D_{A_i,A}}}\right)^2\right)}\sqrt{\left(\sum_{r_j=0}^{\mathrm{R_{max}}}\left(\mathrm{O_{D_{A_i,B}}(r_j)} - \overline{\mathrm{O}}_{\mathrm{D_{A_i,B}}}\right)^2\right)}}$$

where $\mathrm{O_{D_{A_I,A}}}$ is the rank of $\mathrm{D_{A_I,A}}$, and $\overline{\mathrm{O}}_{\mathrm{D_{A_I,A}}}$ is the mean of $\mathrm{O_{D_{A_I,A}}}$. The CBC value of A_i is calculated using

$$C_{A_i} = S_{A_i}\exp\left(-\frac{E_{A_i,B}}{R_{max}}\right)$$

with the $E_{A_i,B}$ as the distance from A_i to the nearest neighbor from species B. CBC values range from –1 to 1.

LaSSI simulations

Simulations were performed using LaSSI, a lattice-based Monte Carlo engine, as described previously [8]. Monte Carlo moves are accepted or rejected based on the Metropolis-Hastings criterion so that the probability of accepting a move is equal to min(1, exp(-$\beta\Delta E$)), where $\beta = 1 / kT$. Here, kT is the simulation temperature and ΔE is the change in total system energy associated with the attempted move. Total system energies were calculated using a nearest-neighbor model and previously derived interaction parameters, which have been shown to accurately capture the phase behavior of A1-LCD and variants thereof[3]. For each simulation, 200 distinct chain molecules comprised of 137 beads each were placed in a cubic lattice with length 120 lattice units. Previous calibrations have shown that the numbers of molecules used in these simulations are adequate to avoid problems due to finite size effects. The simulations were performed for 10^{12} Monte Carlo moves and were allowed to equilibrate such that the chains formed a single condensate with a coexisting dilute phase before any analysis was performed. Given our interest in the dynamics of the condensate chains, we limited the Monte Carlo move set to local moves, eliminating non-physical moves that are part of the total set of moves. The jumping distance of chains in the condensate was calculated by tracking the distance moved by the center of mass of each chain after $2.5*10^8$ total system Monte Carlo moves. The number of sticker-sticker interactions was calculated by counting the number of intermolecular interactions between aromatic beads (Phe or Tyr). Beads are interacting if they are within $\sqrt{3}$ lattice units. We performed five independent simulations, and the results shown aggregate all replicates.

Supplementary Tables

Table S1: Protein sequences used in this study. Sequences for the wild-type hnRNPA1 low-complexity-domain (A1-LCD), A1-LCD with an N-terminal His-tag used for purification, and the RG-rich region from DDX4. The amino acids left of the vertical line in A1-LCD with His-tag are part of the His-tag. Color coding of amino acids is as follows: black – nonpolar (A, C, I, L, M, V), green – polar (G, H, N, S, T, Q), red – acidic (D, E), blue – basic (K, R), yellow – aromatic (F, W, Y), and magenta – proline (P).

A1-LCD	GSMASASSSQ RGRSGSGNFG GGRGGGFGGN DNFGRGGNFS GRGGFGGSRG GGGYGGSGDG YNGFGNDGSN FGGGGSYNDF GNYNNQSSNF GPMKGGNFGG RSSGGSGGGG QYFAKPRNQG GYGGSSSSSS YGSGRRF
A1-LCD with His-tag	MSYYHHHHHH LESTSLYKKA GFENLYFQ|GS MASASSSQRG RSGSGNFGGG RGGGFGGNDN FGRGGNFSGR GGFGGSRGGG GYGGSGDGYN GFGNDGSNFG GGGSYNDFGN YNNQSSNFGP MKGGNFGGRS SGGSGGGGQY FAKPRNQGGY GGSSSSSSYG SGRRF
DDX4-IDR	MGDEDWEAEI NPHMSSYVPI FEKDRYSGEN GDNFNRTPAS SSEMDDGPSR RDHFMKSGFA SGRNFGNRDA GECNKRDNTS TMGGFGVGKS FGNRGFSNSR FEDGDSSGFW RESSNDCEDN PTRNRGFSKR GGYRDGNNSE ASGPYRRGGR GSFRGCRGGF GLGSPNNDLD PDECMQRTGG LFGSRRPVLS GTGNGDTSQS RSGSGSERGG YKGLNEEVIT GSGKNSWKSE AEGGES

Table S2: The concentration of fluorogenic probes for epifluorescence imaging, SM imaging and SM tracking.

Fluorogenic probes	Nile blue	Nile red	Merocyanine 540
Epifluorescence imaging	40 μM	34 μM	17 μM
SM imaging/tracking	2 nM	3.4 μM	1.7 μM

Table S3: Microscope setup for imaging condensates using three fluorogenic probes.

Fluorogenic probes	Nile blue	Nile red	Merocyanine 540
Excitation laser	Coherent OBIS 637 nm	Coherent Sapphire 561 nm	Cobolt 06-DPL 532 nm
Dichroic mirror	Chroma ZT532/640rpc-UF1	Semrock Di01-R488/561	Chroma ZT532/640rpc-UF1
Filter	Semrock FF01-676/37	Semrock FF01-593/46	Semrock FF01-593/46

Table S4: Laser power used for pumping fluorogenic probes.

Fluorogenic probes	Nile blue	Nile red	Merocyanine 540
Epifluorescence imaging	~51 W/cm^2	~51 W/cm^2	~51 W/cm^2
SM imaging/tracking	~4088 W/cm^2	~4088 W/cm^2	~4088 W/cm^2

Table S5: Fluorescence statistics of NB and NR. The number of localizations for each condensate, fluorescence signal per localization, burst duration, and displacement velocity for measurements shown in **Fig.4h** (or **Extended data Fig. 8n**) and **Extended data Fig. 8g**. The NB data were averaged over six condensates and NR data were averaged over eleven condensates.

Fluorogenic probes	Nile blue	Nile red
Number of localizations/condensate	$6.0*10^4$	$2.9*10^4$
Average signal (photons/localization)	149	173
Average burst duration (ms)	41	36
Average displacement velocity (nm/10-ms)	100	75

SI_video_floating_NR_A1_LCD.mp4

SI Video 1 **Hubs revealed by NR within a floating A1-LCD condensate.**

SM_reconstruction frames_resized.mp4

SI Video 2 **Detecting and localizing single NR molecules within an A1-LCD condensate**. Left: Position estimates (green crosses) overlaid on the raw SM images. Right: SMLM reconstruction formed by accumulating localizations over time. Each detected single molecule is represented as a shaded red disk.

NR_sequence_video_labelled.mp4

SI Video 3 **SMLM images of NR within an A1-LCD condensate reconstructed using subsets of raw SM images**. The hubs of NR disappear and appear at nearby positions over time (sec). Yellow circles: hubs.

NB_sequence_video_labelled.mp4

SI Video 4 **SMLM images of NB within an A1-LCD condensate reconstructed using subsets of raw SM images**. The hubs of NB disappear and appear at nearby positions over time (sec). Yellow circles: hubs.

SM_tracking.mp4

SI Video 5 **Single-molecule tracking of NB within A1-LCD condensates**. Yellow crosses represent the position of each SM, while yellow lines represent the trajectory of each SM overlaid on the raw SM images.

Supplementary References

1 Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699, doi:10.1126/science.aaw8653 (2020).

2 Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences USA* **114**, E8194-E8203, doi:doi:10.1073/pnas.1706197114 (2017).

3 Mazidi, H., Lu, J., Nehorai, A. & Lew, M. D. Minimizing Structural Bias in Single-Molecule Super-Resolution Microscopy. *Scientific Reports* **8**, 13133, doi:10.1038/s41598-018-31366-w (2018).

4 Ferdman, B. *et al.* VIPR: vectorial implementation of phase retrieval for fast and accurate microscopic pixel-wise pupil estimation. *Opt. Express* **28**, 10179-10198, doi:10.1364/OE.388248 (2020).

5 Ripley, B. D. Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 172-192, doi:https://doi.org/10.1111/j.2517-6161.1977.tb01615.x (1977).

6 Khater, I. M., Nabi, I. R. & Hamarneh, G. A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods. *Patterns* **1**, doi:10.1016/j.patter.2020.100038 (2020).

7 Malkusch, S. *et al.* Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochemistry and Cell Biology* **137**, 1-10, doi:10.1007/s00418-011-0880-5 (2012).

8 Farag, M. *et al.* Condensates formed by prion-like low-complexity domains have small-world network structures and interfaces defined by expanded conformations. *Nature Communications* **13**, 7722, doi:10.1038/s41467-022-35370-7 (2022).