GelMetrics: An algorithm for analyzing the dynamics of gel-like phase separated condensates

Authors: Naor Granik[1], and Roee Amit[2,3*]

Affiliations:

1 Department of Applied Mathematics, Technion - Israel Institute of Technology, Haifa 32000, Israel.

2 Department of Biotechnology and Food Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.

3 The Russell Berrie Nanotechnology Institute, Technion - Israel Institute of Technology, Haifa 32000, Israel.

* To whom correspondence should be addressed: roeeamit@technion.ac.il.

Abstract

Cellular compartments and organelles are essential for the spatial organization of biological matter. Recently, membraneless organelles like paraspeckles, stress granules, and Cajal bodies have garnered significant scientific interest due to their lack of membrane boundaries and crucial cellular functions. These organelles self-assemble through phase separation, a process in which a homogeneous solution separates into distinct phases. The phases most commonly encountered in cells are liquids and gels. Various microscopy techniques exist to study these phase-separated compartments. However, these are often inadequate for investigating the dynamics of gel-like condensates, where molecular motion occurs over tens of minutes rather than seconds. Here, we introduce a method to quantitatively measure the dynamics of gel-like phase-separated organelles by tracking their fluorescence signals over extended durations. We demonstrate that our algorithm can identify biological activity amidst measurement noise and estimate biophysical parameters which can provide insights into the dynamic behavior of the condensates. We validated our approach on synthetic RNA-protein granules, demonstrating its applicability both in vitro and in vivo.

Background

Cellular compartments and organelles play a crucial role in the structural organization of biological matter. Well-established compartments such as the nucleus or Golgi apparatus are distinguished by the presence of a lipid membrane boundary that separates them from the surrounding cellular environment. In addition to these conventional membrane-bound organelles, there exists another category of cellular compartments known as membraneless organelles. This category includes compartments such as paraspeckles, stress granules and Cajal bodies to name a few [1, 2]. These

cellular entities are unique in that they remain distinct from the cellular environment despite the absence of a membrane, while still performing numerous functions. Recent research suggests that these membraneless organelles, often characterized as supramolecular assemblies of proteins and RNA molecules, self-assemble through the process of phase separation [3–6].

Phase separation is the process by which a homogeneous solution separates into multiple distinct phases, each with unique compositions and properties. Typically, phase separation processes are classified by the different material states of the resulting phases (e.g., liquid, gel, glass), which can lead to various transitions (for example, a liquid-solid transition indicates a transformation from a liquid environment into a solid-like compartment). The forms commonly encountered in cellular biology are liquid-liquid and liquid-gel, however determining the characteristics and driving forces in a phase separating system within a living cellular environment remains a challenging task.

A primary method of investigation of naturally occurring membraneless organelles is in vitro reconstitution. This method entails incubation of individual components of condensates in specific conditions to test whether they phase separate in vitro [7–9]. Alternative methods attempt to probe the dynamics of membraneless organelles using microscopy. These include the commonly used fluorescence recovery after photobleaching (FRAP) which is utilized to assess the liquidity of phase separated condensates, and fluorescence resonance energy transfer (FRET) which can measure interaction kinetics within condensates including binding and unbinding kinetics. More recently, fluorescence correlation spectroscopy (FCS) has been utilized to investigate diffusivity at the level of individual molecules assuming sufficient sparsity [10–12].

Although FRAP is perhaps the most extensively utilized fluorescence-based method, its applicability is limited to scenarios where the condensate is liquid-like, in which case diffusion can be observed and measured. Alternatively, FRAP is useful when attempting to determine the material state itself, as fluorescence recovery, or lack thereof, can differentiate between liquid-like and solid-like condensates. Notably, FRAP is not always suitable when investigating dynamics in condensates with solid-like characteristics, specifically, gel-like, where molecular motion exists but at a significantly slower rate [13–15].

Given the slower molecular dynamics of gel-like condensates, we propose an alternative approach for their study that does not rely on specialized microscopy methods. By tracking condensate fluorescence intensities over extended periods, we can address the issue of low diffusivity in the granules by analyzing the fluorescence signals themselves to extract instances of molecular motion. The challenge of finding a weak or rare signal in a noisy environment is known as rare signal recovery. Basic techniques in this field, such as matched filtering, and Bayesian analysis utilize prior knowledge on the underlying signal to separate it from the noise. More advanced approaches like subspace methods and empirical mode decomposition, provide robust tools for identifying weak

signals. These methods and others were designed to tackle complex signals commonly encountered in communications and remote sensing (e.g., radar), as such they might be less effective in the biological context described here.

In this work, we developed a method to quantitatively measure the dynamics of phase separated gel-like granules by tracking their fluorescence signals over extended periods of time. We demonstrate on synthetic signals that via our algorithm we can distinguish between different types of possible signals (constant, gradually changing, and bursty). Moreover, our algorithm successfully extracts important biophysical features from the signals, even in the presence of measurement noise. We then demonstrate the applicability of our algorithm to study synthetic RNA-protein granules both in vitro and in vivo.

Implementation

Assumptions

We refer here to a scenario of a gel-like phase separated, localized granule, with fluorescently marked components (either the components themselves are fluorescent or are tagged with a fluorescent marker) residing within a dilute solution that can itself contain a low concentration of fluorescent molecules. We assume an experimental setting wherein a field of view containing both highly fluorescent granules and a dark background, is monitored for a total duration (T), with constant intervals (t) between fluorescence intensity measurements.

We assume the total fluorescence of the granule is comprised of three distinct signal processes: granule fluorescence, background fluorescence and noise. Granule fluorescence refers to the sum of fluorescence intensities of the molecules residing within the granule core, with the occasional addition or subtraction of intensity resulting from molecules entering or exiting the granule. Background fluorescence refers to the fluorescence intensity of the dilute phase encircling the bright granule which contains similar components to the granule albeit at a much lower concentration. Noise here refers to various sources of noise which commonly affect fluorescence microscopy, including but not limited to shot noise, electronic noise, and dark current noise [16]. We further assume that background fluorescence is slowly changing, as compared with granule fluorescence which depends on the dynamic insertion and shedding events occurring in the granule. Finally, we consider noise to be a symmetric, memory-less process.

Based on these assumptions, we define a general model for the fluorescence intensity signal comprised of both additive and exponential components, corresponding to fluorescent molecules, and photobleaching. This can be described as follows:

$$y(t) = (S(t) + c(t)) \cdot f(t) \tag{1.1}$$

$$c(t) = c_0(t) \cdot f(t) \tag{1.2}$$

where $y(t)$ is the observed fluorescent signal, $S(t)$ is the underlying granule signal which we try to extract, $c(t)$ is the observed background signal, $c_0(t)$ is the underlying background signal and $f(t)$ is the photobleaching component.

Signal extraction

To find the underlying granule signal $S(t)$, we assume that the background signal is a constant:

$$c_0(t) \approx c_0 = const \tag{1.3}$$

This leads to:

$$\frac{y(t)}{c(t)} = \frac{S(t) + c_0}{c_0} \tag{1.4}$$

$$S(t) = c_0 \left(\frac{y(t)}{c(t)} \right) - c_0 = c_0 \left(\frac{y(t)}{c(t)} - 1 \right) \tag{1.5}$$

Noticeably, this model lacks the contribution of high frequency noise effects stemming from physical measurements in a fluorescence microscope. These affect the measured granule signal $y(t)$, and the measured background signal $c(t)$. To reduce these effects, and since we do not require the background signal at high temporal resolution, we fit the measured background signal to a 3rd degree polynomial. This is done to capture the general trend of the signal (i.e., the photobleaching component) while eliminating fluctuations due to random noise. In contrast, the measured granule signal which does contain biologically relevant data is filtered with a moving average filter with a window span dependent on experimental considerations.

Identifying signal bursts

A hallmark of phase separated compartments is the ability of macromolecules to traverse the phase boundary in either direction [17, 18]. In terms of fluorescence intensity, such movement should ideally manifest as a change or shift in the level of signal intensity, lasting over multiple time points and leading to a higher or lower sustainable signal intensity level. We henceforth term such shifts in the signal as “signal bursts”. To identify such shifts in the base-line fluorescence intensity, we apply a moving-average filter to smooth the data (note that this is the same filtering step as described in the signal extraction section). The effect of such an operation is to bias the fluctuations of the smoothed noisy signal in the immediate vicinity of the bursts towards either a gradual increase or decrease in the

signal (Figure 1a). Random single fluctuations, which do not settle on a new baseline level are not expected to generate a gradual and continuous increase or decrease over multiple time-points in a smoothed signal. We note that this operation alone can mask biologically relevant events that occur on very short timescales, rendering this algorithm less effective in the case of a liquid-like condensate with fast molecular motion.

Following this filtering step, we search for contiguous segments of gradual increase or decrease and record only those whose probability for occurrence is 1 in 1000 or less given a Null hypothesis of randomly fluctuating noise. To translate this probability to a computational threshold, we first compute the intensity difference distribution for every trace separately. This distribution is computed by collecting all the instantaneous differences in signal $\Delta S(t_i) = S(t_i) - S(t_{i-1})$ and binning them (Figure 1b). Given a particular trace the likelihood for observing an instantaneous signal increase event in a time-point (t_i) can therefore be computed as follows:

$$P_{inc} = \frac{N(\Delta S(t_i) < 0)}{N_{tot}} \tag{1.6}$$

Where $N(\Delta S(t_i) > 0)$ and N_{tot} correspond to the number of increasing instantaneous events and total number of events in a trace respectively. Likewise, the number of decreasing instantaneous events is defined as:

$$P_{dec} = \frac{N(\Delta S(t_i) < 0)}{N_{tot}} \tag{1.7}$$

This in turn allows us to compute the number of consecutive instantaneous signal increase events (m) to satisfy our 1 in 1000 or less threshold for a significant signal increase burst event m as follows:

$$P_{inc}^{m} = \frac{1}{2^j} \Rightarrow m \log_2(p_{inc}) = -j \Rightarrow m = -\frac{j}{\log_2(p_{inc})} \tag{1.8}$$

Here, (j) is a stringency parameter which allows us to control the threshold probability. For example, a value of $j = 10$ would correspond to a probability of 1/1024 which is the maximal probability we allow. This freedom of stringency is required since different systems will have different levels of noise, resulting in more false positive events being detected. The threshold is calculated for each signal separately and an analogous threshold is calculated for decrements in the signal and is typically in the range $[m - 1, m + 1]$.

Using this definition of m, we can identify segments within the signal trace corresponding to signal bursts. Segments within the signal that are not classified as either a negative or positive burst event are considered unclassified. Unclassified segments are typically signal elements whose noise profile

does not allow us to make a classification into one or the other event-type. In Figure 1c we mark the classifications on a sample trace with positive bursts, negative burst, and non-classified events in green, red, and blue, respectively. We confine our segment analysis between the first and last significant segments identified in each signal, since we cannot correctly classify signal sections that extend beyond the observed trace. For each identified segment we record the amplitude (ΔI) and time between successive events of the same type (e.g., the time between two positive burst events).

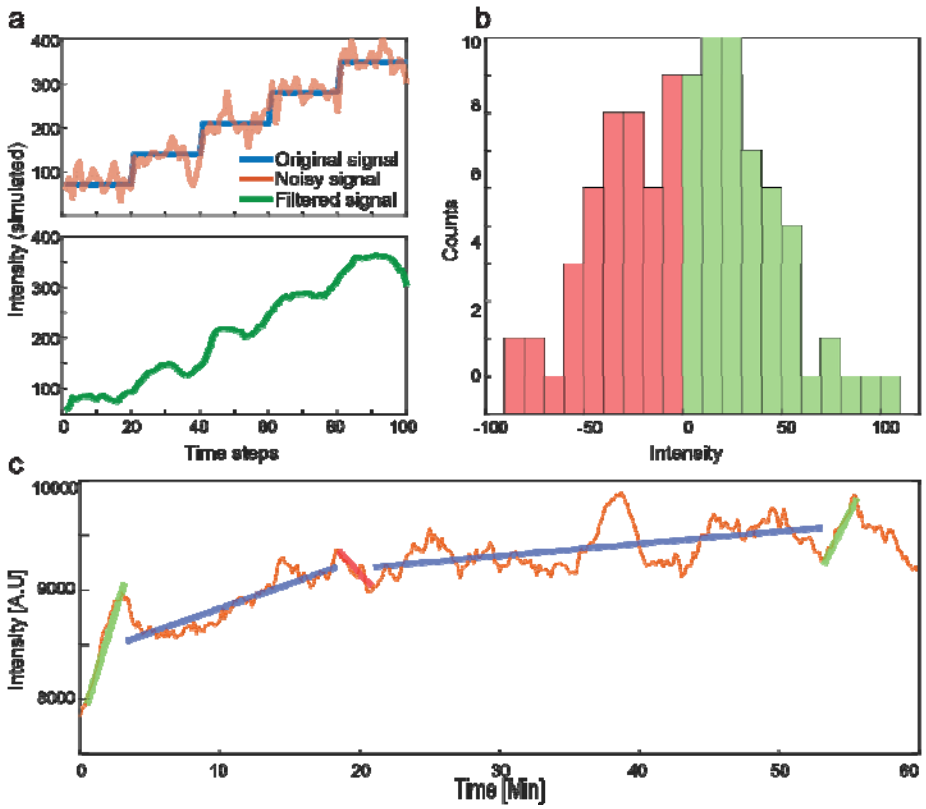


Figure 1 – Identification of burst events.

a. Top: simulated step signal (blue) with added white Gaussian noise (orange). Bottom: noisy signal after moving average filter. **b.** Intensity difference distribution for the signal presented in panel A. **c.** Sample experimental signal (orange) overlaid with markers indicating identified segments in green, blue, and red, corresponding to positive bursts, quiescent segments, and negative bursts.

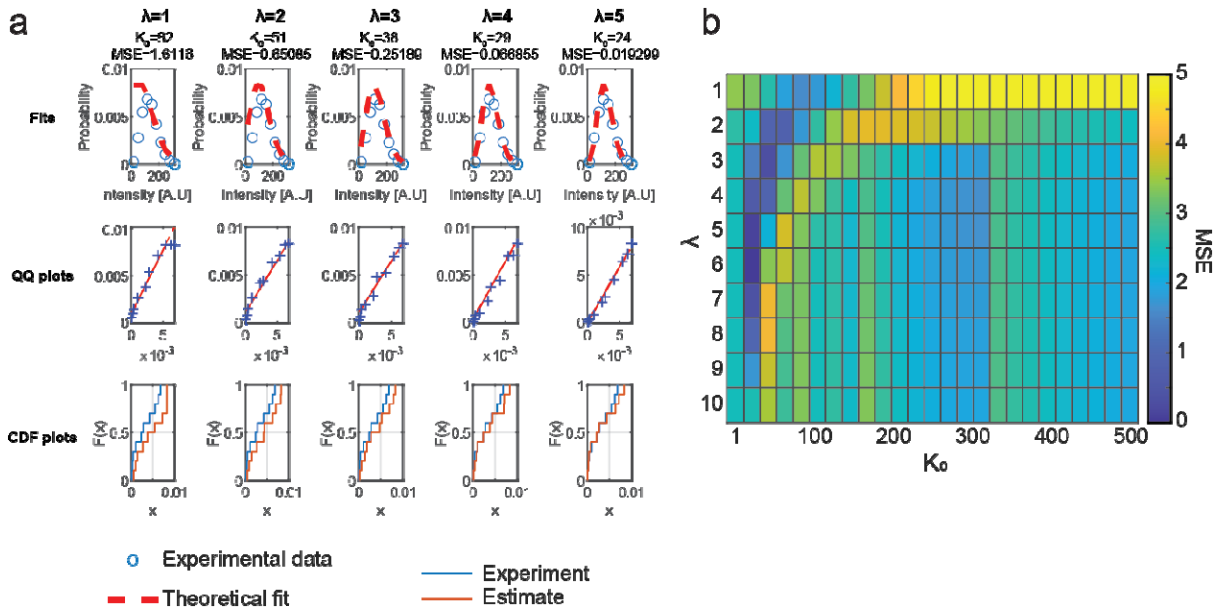
Estimating signal parameters

Our model posits that a varying number of discrete molecules participate in a burst event (either positive or negative), therefore a Poisson distribution would be best suited for describing the data.

However, given the fact that we cannot directly infer the fluorescence intensity of a single macromolecule, we chose to fit the amplitude distribution to a modified Poisson function of the form:

$$p(I) = \frac{\lambda^{\frac{I}{k_0}} e^{-\lambda}}{\left(\frac{I}{k_0}\right)!} \tag{1.9}$$

where I is the observed fluorescence amplitude, λ is the Poisson parameter (rate), and is a fitting parameter whose value corresponds to the fluorescence intensity associated with a single fluorescent molecule within the burst. For each pair of λ and values we calculate the theoretical Poisson distribution (Figure 2a) and evaluate its deviation from the experimental amplitude distribution in terms of mean squared error (MSE). We then present the resulting estimates for λ=1,..,5, together with



the corresponding empirical QQ-plots as a qualitative metric of goodness of fit. (Figure 2b)

Materials and methods

Formation of in vitro RNA-protein granules

Construction of the required plasmids for RNA and protein expression, and expression/synthesis of the components was done as previously described in [18]. Briefly, a vector containing the DNA sequence encoding for the required RNA was linearized via digestion with a restriction enzyme, followed by an in vitro transcription reaction with a HiScribe T7 RNA synthesis kit (New England

Figure 2 – Poisson fitting example results.

a. Sample fitting of an amplitude distribution to five modified Poisson functions, each with a different

Biolabs (NEB) #E2040). Following in vitro transcription, the remaining DNA in the reaction was degraded using a DNAse I enzyme (NEB #M0303S). RNA products were purified using Monarch RNA Cleanup Kit (NEB, #T2040S) and immediately used or stored in −80° for later use.

For protein purification, E. coli cells expressing the fusion protein (PP7 coat protein-mCherry were grown overnight with appropriate antibiotics at 37 °C with 250 rpm shaking, and diluted the following morning 1/100 into Terrific Broth (TB: 24 g yeast extract, 20 g tryptone, 4 ml glycerol in 1 L of water, autoclaved, and supplemented with 17 mM KH2PO4 and 72 mM K2HPO4), with appropriate antibiotics and induction (500 µl C4-HSL) and grown in 37 °C and 250 rpm shaking to optical density (OD) > 10. Cells were harvested, resuspended in 30 ml resuspension buffer (50 mM Tris-HCl pH 7.0, 100 mM NaCl and 0.02% NaN3), disrupted by four passages through an EmulsiFlex-C3 homogenizer (Avestin Inc., Ottawa, Canada), and centrifuged (10,000 g for 30 min) to obtain a soluble extract. Fusion protein was purified using HisLink Protein purification resin (Promega (Madison, WI) #V8821) according to the manufacturer's instructions. Buffer was changed to 1xPBS (Biological Industries, Israel) using multiple washes on Amicon ultra-2 columns (Merck, Burlington, MA #UFC203024).

In vitro experiments were performed in granule buffer (final concentrations: 750 mM NaCl, 1 mM MgCl2, 10% PEG4000). Reactions were set up at the appropriate concentrations and allowed to rest at room temperature for 1 h. 3–5 µl from the reaction was then deposited on a glass slide prior to microcopy.

Formation of in vivo RNA-protein granules

BL21-DE3 cells expressing the two plasmid system (i.e., one plasmid encoding for the RNA binding sites array, and a second plasmid containing the fluorescent protein fused to a phage coat protein) were grown overnight in 5 ml Luria Broth (LB), at 37° with appropriate antibiotics (Cm, Amp), and in the presence of two inducers – 1.6 µl Isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration 1 mM), and 2.5 µl C4-HSL (final concentration 60 µM) to induce expression of T7 RNA polymerase and the RBP-FP, respectively. Overnight culture was diluted 1:50 into 3 ml semi-poor medium consisting of 95% bioassay buffer (BA: for 1 L—0.5 g Tryptone [Bacto], 0.3 ml glycerol, 5.8 g NaCl, 50 ml 1 M MgSO4, 1 ml 10×PBS buffer pH 7.4, 950 ml DDW) and 5% LB with appropriate antibiotics and induced with 1 µl IPTG (final concentration 1 mM) and 1.5 µl C4-HSL (final concentration 60 µM). The liquid culture was shaken for 3 h at 37° before being applied to a gel slide [3 ml PBSx1, mixed with 0.045 g SeaPlaque low melting Agarose (Lonza, Switzerland), heated for 20 seconds and allowed to cool for 25 min]. 1.5 µl cell culture was deposited on a gel slide and allowed to settle for an additional 30 min before imaging.

Fluorescence Microscopy

In vitro granules and live E. coli cells were imaged in a Nikon Eclipse Ti-E epifluorescent microscope (Nikon, Japan) with a 100×1.45 NA oil immersion objective. A time lapse experiment was carried out by tracking a field of view for 60 min on Nikon Eclipse Ti-E epifluorescent microscope (Nikon, Japan) using an Andor iXon Ultra EMCCD camera at 6 frames-per-minute with a 250 msec exposure time per frame. Excitation was performed by a CooLED (Andover, UK) PE excitation system at 585 nm for the mCherry protein. Microscopy control and data acquisition was performed using Nikon NIS-Elements version 4.20.02 (build 988) 64 bit.

Image analysis and signal extrapolation

The brightest spots (top 10%) in the field of view were tracked over time and space using the imageJ MosaicSuite plugin [19–21]. A field of view typically contained dozens of granules (in-vitro) or cells containing puncta (in vivo).

The tracking data, (x,y,t coordinates of the bright spots centroids), together with the raw microscopy images were processed by a custom built Matlab (The Mathworks, Natick, MA) script designed to normalize the relevant spot data according to the described model. The normalization involved the following steps: for each bright spot, a 14-pixel wide sub-frame was extracted from the field of view, with the bright spot at its center. Each pixel in the sub-frame was categorized based on its mean intensity value to one of three categories. The brightest pixels were classified as ‘spot region’, typically appearing as a cluster which corresponds to the spot itself. The dimmest pixels were classified as ‘dark background’, representing empty regions in the field of view. Lastly, values in between were classified as ‘cell background’. For in vitro experiments the ‘dark background’ and ‘cell background’ pixel groups yield similar intensity values, however, this does not affect performance. From each sub-frame, two values were extracted: the mean of the ‘spot region’ pixels and the mean of the ‘cell background’ pixels, corresponding to spot intensity and background intensity values. This was repeated for each spot from each frame in the data, resulting in sequences of intensity vs. time for both the spot and cell background.

Results

Differentiating between different sets of simulated data

As an initial evaluation of our algorithm, we simulated three distinct types of base signals. For each simulation category, we generated and analyzed 1000 signals consisting of 360 time-points to align with our experimental setup. The three signal types are: (1) signals featuring multiple burst events aimed at emulating the characteristic behavior of phase-separated condensates, reflecting molecules transitioning between dense and dilute phases; (2) constant flat signals representing a stationary

observable akin to a phase-separated solid, (e.g., a glass phase with arrested dynamics); and (3) gradually increasing signals simulating a continuous flux of fluorescent molecules moving unidirectionally. In all simulations, we incorporated two noise components according to our noise model: white Gaussian noise with a peak-to-peak amplitude of 40 [A.U], matching the value estimated from experimental traces in our hands, and an exponential component simulating photobleaching (Figure 3).

Prior to analysis, we first calibrated our algorithm such that the constant signals will output roughly one identified event per 1000 time points. For this, we applied our burst detection algorithm on the flat constant signals with three different levels of stringency (parameter *j* in equation 1.8) – 10,12,14 corresponding to probabilities of $\frac{1}{1024}, \frac{1}{4096}, \frac{1}{12384}$. In accordance with the given parameters, our algorithm identified 5.2, 2.5, and 1.1 segments per 1000 time points. Thus, a threshold parameter of 14, corresponding to a probability of $\frac{1}{12384}$ was determined to be optimal for the simulation parameters stated above. We note that our coarse simulation does not perfectly represent biological signals with all their innate complexity, therefore this setting might not be optimal for an experimental signal.

We first applied our burst-detection algorithm to 1000 signals containing randomly distributed, increasing, or decreasing instantaneous bursts with an amplitude of $60\ [\mathrm{A.U}] \times \mathrm{m}$ with m being a random number generated from a Poisson distribution with λ=2. (Figure 3a) The resulting burst amplitudes lie in the range 0-600 [A.U] and appear symmetric in nature with relatively similar numbers of positive and negative bursts detected (>1000) (Figure 3b). Analysis of the constant simulated signals (Figure 3c) reveals once again a symmetric positive and negative amplitude distribution; however, a close examination reveals that the burst amplitude width is smaller by a factor of ~5 compared to the results of the bursty signals. To compute whether the number of burst events identified via our algorithm is statistically significant, we utilized a chi-squared test. Our algorithm identified 414 positive bursts in 1000 constant signals. For 1000 simulated signals containing bursts, 1720 positive bursts were identified, yielding a p-value of 1e-5 indicating a statistically significant difference between constant and bursty signals. In addition, the total number of events observed is ~300 for both positive and negative events, which indicates roughly one event per 1000 time points, in accordance with our base assumption (Figure 3d). For the gradually increasing signals (Figure 3e, f) a negligible number of negative burst-like events was detected by our algorithm, with a pronounced bias towards positive events (<10 negative events vs. ~500 positive events). The scarcity of events can be explained by the positive bias in the signal which results in a steep increase in the statistical threshold for event identification. Similar simulations with a decreasing signal show a mirror image of amplitude distribution (data not shown). Taken together, these results demonstrate that through our

algorithm, we can differentiate between different types of signals, each corresponding to a different underlying behavior profile.

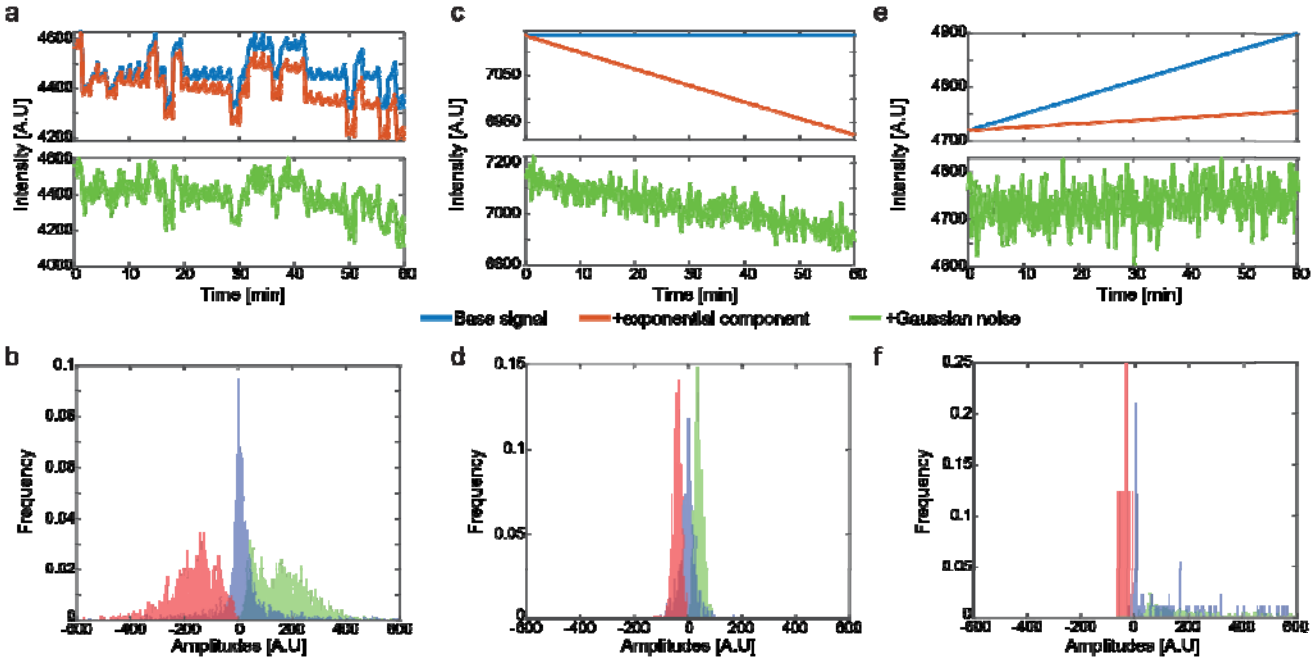


Figure 3. Differentiating between different types of signals.

a. Simulated signal with burst events (blue), with an added photobleaching component (orange), and added Gaussian noise component (green). **b.** Amplitude distributions of burst events identified from 1000 signals containing intermittent bursts. **c.** Simulated constant signal (blue), with an added photobleaching component (orange), and added Gaussian noise component (green). **d.** Amplitude distributions of burst events identified from 1000 constant signals. **e.** Simulated gradually ascending signal (blue), with an added photobleaching component (orange), and added Gaussian noise component (green). **f.** Amplitude distributions of burst events identified from 1000 sloped signals.

Estimating simulated signal parameters

The simulated bursty signals were generated with burst amplitudes which are a product of two values: a constant intensity , and a random number generated from on an underlying Poisson distribution with a specified λ parameter. To examine whether we can correctly identify the signal parameters, we generated three sets of simulated signals with different values - 1,2,4, and fitted the extracted positive amplitude data to our modified Poisson function.

Figure 4 displays the positive amplitude distributions for all three cases, demonstrating a visible change in the distributions with higher values (Figure 4a, b, c). The estimation results corresponding to the λ values used in the simulation are also presented (Figure 4d, e, f). In all three

scenarios, the algorithm managed to estimate a k_0 value relatively close to the ground truth (60 [A.U]). $k_{0_{est}} = 68$ for $\lambda_{est} = 1$, 67 for $\lambda_{est} = 2$, and 63 for $\lambda_{est} = 4$. Similarly, the algorithm estimated λ_{est} values within the range $[\lambda_{sim} - 1, \lambda_{sim} + 1]$. For $\lambda_{sim} = 1$ the best estimate according to the MSE metric was $\lambda_{est} = 2$, For $\lambda_{sim} = 2, \lambda_{est} = 2$, while for signals generated with $\lambda_{sim} = 4$, the best estimate was $\lambda_{est} = 3$. This slight discrepancy, underscores the importance of verifying λ_{est} values through visual inspection of the QQ-plots (e.g., for the $\lambda_{sim} = 1$ case, inspection of the QQ-plot would indicate a better fit for $\lambda_{est} = 1$ than for 2), or by additional controls and experiments.

We further tested the estimation performance under different signal to noise ratios (SNRs) by simulating two more sets of signals. The first set with $\lambda_{sim} = 2$ and $k_0 = 200$, and a white Gaussian noise of magnitude 40 [A.U] (SNR of 5). The second set had $\lambda_{sim} = 2$, $k_0 = 60$, and white Gaussian noise of 120 [A.U] (SNR of 0.5). We estimated the Poisson parameters of the amplitude data gathered from these signals (Figure 4g, h). In the high SNR case, our algorithm correctly estimated $\lambda_{est} = 2$, $k_{0_{est}} = 223$ (Figure 4i), In the low SNR case, while k_0 was correctly estimated ($k_{0_{est}} = 61$), the λ value was shifted to $\lambda_{est} = 3$ (Figure 4j). These findings illustrate that high noise levels compared to the base amplitude of the bursts can lead to incorrect estimation of the underlying distribution, with lower signal to noise ratio likely leading to a higher estimation bias.

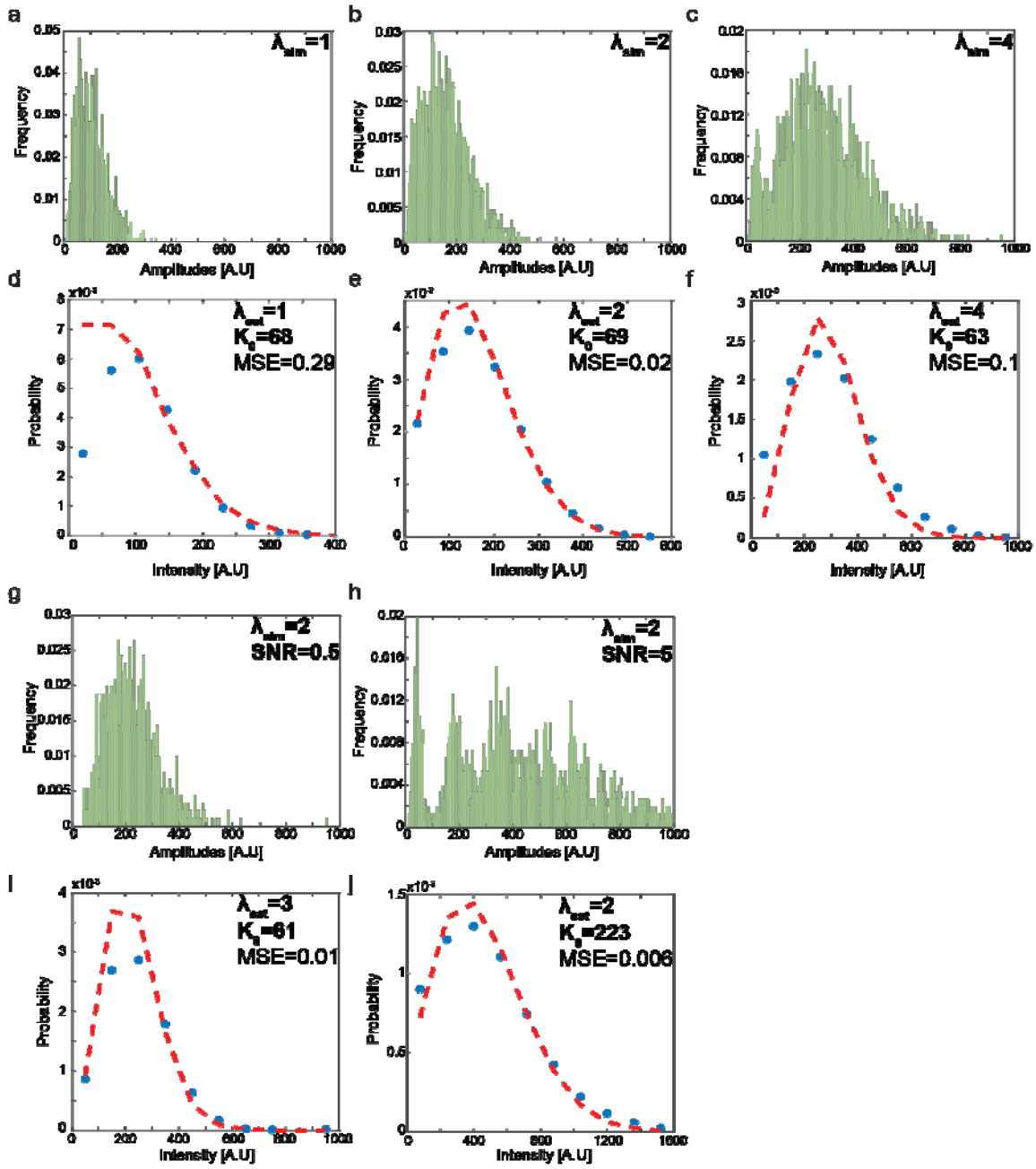


Figure 4. Estimating simulation parameters

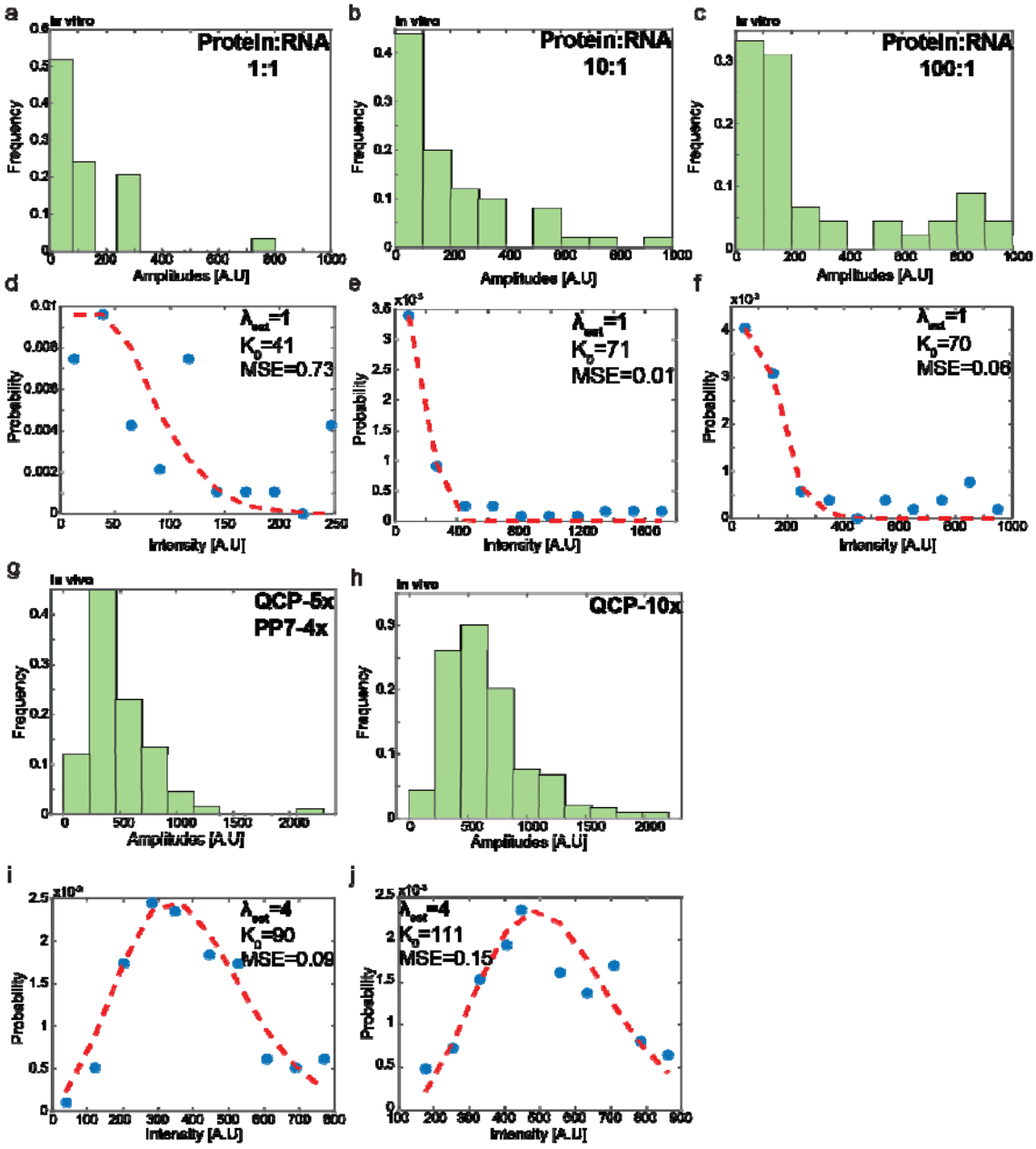
a-c. Amplitude distributions gathered from 1000 bursty simulated signals with similar base intensity (and different Poisson parameter ( ) values. (a) (b) (c) The distributions demonstrate a visible shift in accordance with the Poisson parameter used in the simulation. **d-f.** Poisson distribution fits of the value used in the simulation. Blue points show the measured amplitude distribution, red lines are the theoretical Poisson distribution. The fits show a relative accurate estimation of the base intensity value ( ). **g-h.** Amplitude distributions gathered from 1000 bursty simulated signals with different SNR conditions. (g) SNR=0.5, (h) SNR=5. **i-j.** Poisson distribution fits of the best estimates according to the algorithm in terms of means squared error. In both cases the algorithm correctly estimates the base fluorescence ( ), however in low SNR conditions, the estimation of the Poisson parameter is shifted to

Validation on experimental data

We validated our algorithm on experimental data from RNA-protein (RNP) granules formed both in vitro and in vivo. The granules were produced using our previously established phase separation system, which consists of RNA molecules that fold into secondary structures with multiple hairpins, together with phage coat proteins capable of binding these hairpins. RNA molecules, bound by one or more proteins, intermittently enter, or exit the granules, causing fluctuations in fluorescence levels with the amplitudes of these shifts being dependent on the number of proteins bound to the RNA [18]. We prepared three sets of granules in vitro, with a constant RNA concentration (120 nM) and varying protein concentrations, resulting in 1:1, 10:1 and 100:1 protein to RNA ratios (see methods). The RNA construct, designated as PCP-14x/MCP-15x contains 14 hairpin binding sites for the PP7 phage coat protein (PCP), interspaced by 15 hairpin binding sites for the MS2 phage coat protein (MCP), which serve as structural spacers and are not recognized by PCP [22]. This experimental setup enabled us to establish a scale for evaluating the algorithm's performance, as we anticipated a similar Poisson distribution parameter but different k_0 values depending on the protein content in the system. We tracked the fluorescence intensity of multiple granules over a span of 60 minutes, with measurements taken every 10 seconds, and applied our algorithm. The results show that regardless of protein content, most burst amplitudes appear in the lower end of the spectrum (0-100 [A.U]), suggesting that free floating proteins or RNA molecules bound by a few proteins can traverse the phase boundary more easily (Figure 5a, b, c). We estimated the Poisson parameters of the negative amplitudes, as these represent molecules which exit the granule and therefore are affected by its internal dynamics. In addition, we filtered the amplitudes to remove the top 10% of the data as these would often be outliers which originate from technical difficulties in the granule tracking process. In terms of MSE, the algorithm estimated a λ_{est} value of one for all three experimental cases. The k_0 increased from 41 [A.U] for granules formed with equal concentrations of RNA and protein to 71 [A.U] for granules formed with 10:1 protein to RNA ratio. Interestingly, granules formed with a 100:1 protein to RNA ratio were estimated to have a k_0 value of 70 [A.U], suggesting that the RNA molecules might be at full capacity in terms of protein binding, reaching peak fluorescence at the 10:1 ratio (Figure 5d, e, f).

We further tested our algorithm on experimental data from granules formed in E. coli cells. For this, we utilized the Qβ phage coat protein (QCP) together with two synthetic RNA sequences, one which encodes for 10 hairpin binding sites (designated QCP-10x), and another which encodes for 5 Qβ binding sites and 4 PP7 binding sites which are used as structural spacers (designated QCP-5x/PP7-4x). Plasmids encoding these RNAs and QCP fused to mCherry were co-transformed into E. coli cells, which were grown overnight with full induction of the components. Similarly to the in vitro experiments, we tracked the fluorescence intensity of the granules over 60 minutes, with measurements taken every 10 seconds, and applied our algorithm. The amplitude distributions in this

case present a profile more fitting to a Poisson distribution with and appear to be rather similar in range, suggesting that QCP-10x RNA molecules which are not fully occupied more easily traverse the phase boundary (Figure 5g, h). Analysis of the Poisson estimations (Figure 5i, j) reveal that in terms of MSE, the distributions fit a Poisson with and for QCP-5x/PP7-4x, and QCP-10x, respectively. Taken together, these validation experiments demonstrate that our



algorithm can accurately estimate relevant biophysical parameters which in turn provide insight into the behaviour of gel-like condensates.

Figure 5. Analysis of experimental data

a-c. Amplitude distributions gathered from RNP granules formed in vitro. In all three cases, the granules contain similar RNA concentration but different protein concentration (a) 1:1 protein to RNA ratio, (a) 10:1 protein to RNA ratio, (c) 100:1 protein to RNA ratio, yielding a scale by which we can measure the performance of the algorithm. The amplitude distributions demonstrate a bias to the lower end of the scale, indicating that RNA molecules bound by few proteins can more easily traverse the phase boundary. **d-f.** Poisson distribution fits of the best estimate according to the mean squared error. Blue points show the measured amplitude distribution, red lines are the theoretical Poisson distribution. The fits show similar Poisson behavior ($\lambda = 1$), and an increase in base intensity (k_0) with more protein content when protein content in increased 10-fold. However, granules formed with 100x more protein than RNA show similar base intensity as granules formed with 10x more protein, suggesting saturation of the protein binding capacity of the RNA. **g-h.** Amplitude distributions gathered from RNP granules formed in vivo using two different RNA sequences (a) QCP-5x/PCP-4x which has 5 binding sites for QCP, (b) QCP-10x which has 10 binding sites for QCP. **i-j.** Poisson distribution fits of the best estimate according to the mean squared error. Here the distribution best fit a Poisson function with $\lambda = 4$ and with somewhat close k_0 values, indicating that the QCP-10x RNA is likely not fully occupied by proteins.

Discussion

In this work we describe and assess a method for investigating the dynamics of solid-like phase separated condensate. Various experimental assays are available to researchers studying phase separation and membraneless organelles. However, these are often inadequate for the case of solid-like condensates where dynamic motion exists at a much slower rate in comparison to liquid-like bodies. Our algorithm leverages straightforward fluorescence microscopy measurements of the condensates and employs statistical approaches to identify signal segments likely corresponding to molecules entering or exiting the condensate. By aggregating the amplitudes of these segments collected from multiple granules, we derive a distribution, which we fit to a modified Poisson function characterized by two parameters: λ, representing the average number of molecules traversing the phase boundary, and k_0, denoting the average fluorescence intensity of a single molecule.

We validated the performance of our algorithm using two sets of experimentally formed granules. The first set consisted of granules formed in vitro under controlled conditions. In these experiments, the concentration of the phase-separating agent (RNA) was kept constant, while the concentration of the fluorescent protein varied to create a gradient. Analysis of the signals recorded from these granules revealed that their dynamics correspond to a Poisson distribution with $\lambda = 1$, regardless of protein concentration, indicating an average of one RNA-protein molecular complex traversing the phase boundary during a burst event. In contrast, the fluorescence of a single emitter (k_0) increased in

granules formed with higher protein content (10x) but remained constant in granules prepared with higher protein concentration (100x), suggesting saturation of the protein binding capacity of the RNA molecules. Both these observations demonstrate that our algorithm can detect changes in the composition of the granules and occupancy level of the molecular complexes entering and exiting the granule.

The second set of experiments involved granules formed in vivo. For this, we cloned the necessary components into E. coli cells and tracked the fluorescence of the bright spots that formed in the cells. Here, granule dynamics fit a Poisson distribution with $\lambda = 4$. This contrast with the in vitro results suggests that our synthetic granules are more liquid-like or alternatively more brittle in vivo, allowing multiple molecular complexes to traverse the phase boundary during a single burst event. These findings illustrate the algorithm's ability to distinguish and quantify the dynamic behavior of phase-separated granules in different environments.

The presented algorithm has ample room for further development, for example incorporating machine learning methods to better identify signal segments corresponding to biological events, and adding regularization to the Poisson estimation process to place selective emphasis on the tail end of the distribution which we noticed might bias the results. We expect this method will be useful in cases where different sets of granules need to be compared to one another. A scenario which is generally encountered in perturbation experiments (mutations/knockouts of components) and in in-vitro reconstitution studies where components can be selectively added or removed to probe behaviour. Finally, while initially designed for solid-like condensates, we anticipate that with sufficient temporal resolution and adjustment of internal parameters, the algorithm could also be applied to liquid-like condensates, adding to the vast array of methods that exist today.

Conclusion

In conclusion, we present here a method of investigating non-liquid like phase separated condensates using standard fluorescence microscopy and achieving single molecule resolution. The reported parameters (λ_{sim}, k_0) provide a more nuanced picture of the dynamic nature of such condensates and can be utilized in cases in which researchers want to probe the behaviour of condensates under different conditions (e.g., stress conditions), or compositions (e.g., knockouts or mutations to natural components).

Data Availability

The complete code as well as the data required to recreate the figures in this paper is available in the gelMetrics repository: https://github.com/naorgk/gelMetrics

Competing interests

The authors declare that they have no competing interests.

Funding

This work was greatly supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 851615. This study was partially supported by Israel Innovation Authority.

Contributions

N.G. and R.A. conceived the research and wrote the paper; N.G. implemented the software, performed the experiments and data analysis.

Acknowledgements

Not applicable.

References

1. Hirose T, Ninomiya K, Nakagawa S, Yamazaki T. A guide to membraneless organelles and their various roles in gene regulation. Nat Rev Mol Cell Biol. 2023;24:288–304.

2. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017;18:285–98.

3. Jin X, Lee J-E, Schaefer C, Luo X, Wollman AJM, Payne-Dwyer AL, et al. Membraneless organelles formed by liquid-liquid phase separation increase bacterial fitness. Science Advances. 2021;7:eabh2929.

4. Gomes E, Shorter J. The molecular language of membraneless organelles. J Biol Chem. 2019;294:7115–27.

5. Zhao YG, Zhang H. Phase Separation in Membrane Biology: The Interplay between Membrane-Bound Organelles and Membraneless Condensates. Developmental Cell. 2020;55:30–44.

6. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. Cell Communication and Signaling. 2016;14:1.

7. Alberti S, Saha S, Woodruff JB, Franzmann TM, Wang J, Hyman AA. A User's Guide for Phase Separation Assays with Purified Proteins. J Mol Biol. 2018;430:4806–20.

8. Zhang Y, Shen L. An in vitro Assay to Probe the Formation of Biomolecular Condensates. Bio Protoc. 2023;13:e4813.

9. Ganser LR, Myong S. Methods to Study Phase-Separated Condensates and the Underlying Molecular Interactions. Trends in Biochemical Sciences. 2020;45:1004–5.

10. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell. 2019;176:419–34.

467 11. Bracha D, Walls MT, Brangwynne CP. Probing and engineering liquid-phase organelles. Nat
468 Biotechnol. 2019;37:1435–45.

469 12. Rhine K, Skanchy S, Myong S. Single-molecule and ensemble methods to probe RNP nucleation
470 and condensate properties. Methods. 2022;197:74–81.

471 13. Taylor NO, Wei M-T, Stone HA, Brangwynne CP. Quantifying Dynamics in Phase-Separated
472 Condensates Using Fluorescence Recovery after Photobleaching. Biophysical Journal.
473 2019;117:1285–300.

474 14. Wang Z, Lou J, Zhang H. Essence determines phenomenon: Assaying the material properties of
475 biological condensates. J Biol Chem. 2022;298:101782.

476 15. Alshareedah I, Kaur T, Banerjee PR. Methods for Characterizing the Material Properties of
477 Biomolecular Condensates. Methods Enzymol. 2021;646:143–83.

478 16. Ogama T. A beginner's guide to improving image acquisition in fluorescence microscopy. The
479 Biochemist. 2020;42:22–7.

480 17. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science.
481 2017;357:eaaf4382.

482 18. Granik N, Katz N, Willinger O, Goldberg S, Amit R. Formation of synthetic RNA protein granules
483 using engineered phage-coat-protein -RNA complexes. Nat Commun. 2022;13:6811.

484 19. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat
485 Methods. 2012;9:671–5.

486 20. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-
487 source platform for biological-image analysis. Nat Methods. 2012;9:676–82.

488 21. Sbalzarini IF, Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in
489 cell biology. Journal of Structural Biology. 2005;151:182–95.

490 22. Katz N, Tripto E, Granik N, Goldberg S, Atar O, Yakhini Z, et al. Overcoming the design, build, test
491 (DBT) bottleneck for synthesis of nonrepetitive protein-RNA binding cassettes for RNA applications.
492 Nature Communications. 2021.

493

494

495

496