

Characterization and applications of synthetic RNA-protein granules

Research Thesis

**In Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy**

Naor Granik

Submitted to the Senate of the Technion – Israel Institute of Technology

Tammuz 5784

Haifa

July 2024

This Research Thesis was done in the Interdepartmental Program of Applied Mathematics Under the Supervision Prof. Roei Amit.

The Generous Financial Help of the Gutwirth Foundation and of the Technion, Israel Institute of Technology is Gratefully Acknowledged.

The author of this thesis states that the research, including the collection, processing, and presentation of data, addressing, and comparing to previous research, etc., was done entirely in an honest way, as expected from scientific research that is conducted according to the ethical standards of the academic world. Also, reporting the research and its results in this thesis was done in an honest and complete manner, according to the same standards.

Publications

Granik, Naor, Noa Katz, Or Willinger, Sarah Goldberg, and Roe Amit. "Formation of synthetic RNA Protein granules using engineered phage-coat-protein-RNA complexes". *Nature Communications* (2022) (Related to chapters 1,2,3 of this thesis).

Granik, Naor, Roe Amit. "GelMetrics: An algorithm for analyzing the dynamics of gel-like phase separated condensates". *BMC Bioinformatics* (in review) (Related to chapter 2 of this thesis).

Granik, Naor, Sarah Goldberg, and Roe Amit. "Formation of multiphase RNP granules by intrinsically disordered Q β coat proteins and hairpin-containing RNA". *Nucleic Acid Research* (in review) (Related to chapter 4 of this thesis).

Table of Contents

Abstract.....	1
Abbreviations and Notations	2
Introduction	3
Background	6
Phase separation.....	6
Gelation.....	7
Importance of phase transitions in cells.....	7
Polyphasic linkage theory	8
Intrinsically disordered protein regions.....	9
RNA phase transitions in repeat expansion disorders.....	10
RNA stem-loops	10
Phage coat proteins	12
Research Objectives.....	13
Goals	13
Significance	13
Materials and Methods.....	13
Results.....	22
Chapter 1 – Formation of synthetic RNA and RNA-protein granules	22
Hairpin containing RNA phase separates in vitro into gel-like granules.....	22
RNA-based granules co-localize with protein-binding partners.....	27
Chapter 2 – Development of an analysis algorithm for gel-like granules.....	32
Assumptions.....	32
Signal extraction.....	32
Identifying signal bursts	33
Estimating signal parameters.....	35
Estimating simulated signal parameters.....	38
Chapter 3 – Characterization of synthetic gel-like granules	41
Dynamic signal analysis of SRNP granules reveals a structure with gel-like characteristics	41
Competition experiments reveal slow granule equilibration times	44
SRNP granules function as protein capacitors	46
Expression of slncRNAs and protein in bacteria yields puncta-like condensates.....	47
slncRNA expression increases cellular protein concentration.....	50
Chapter 4 – Characterization of gel-like granules with an intrinsically disordered region.....	53

RNA with valency of $n=3$ can melt Q β coat protein condensates within bacterial cells	53
Synthetic RNP granules with Q β coat protein exhibit multiphasic behavior	54
QCP-mCherry titer appears to exhibit a non-monotonic dependence on slncRNA valency	58
Multiphasic Q β granules can control protein titers for genes encoded on the slncRNA	60
Discussion.....	63
References	67

List of Figures

Figure 1 – Theoretical free energy and chemical potential curves.....	6
Figure 2 - Membraneless organelles in the cell.	8
Figure 3 - Schematic of polyphasic linkage	9
Figure 4 - RNA tagging using phage coat proteins.....	12
Figure 5 - In vitro RNA-only granules microscopy.....	23
Figure 6 - Further RNA-only granule analysis.	25
Figure 7 – In vitro RNA-only granules Poisson analysis.	27
Figure 8 - In vitro RNA-Protein granule microscopy.	29
Figure 9 - RNA-Protein granules intensity analysis.	31
Figure 10 - Identification of burst events.....	35
Figure 11 - Poisson fitting example.....	36
Figure 12- Signal type simulations.	38
Figure 13 - Estimating simulation parameters.....	40
Figure 14 - In vitro dynamic signal analysis.....	43
Figure 15 - Competition experiments provide support for liquid-gel phase transition.....	45
Figure 16 - Granules act as protein capacitors.	47
Figure 17 - Synthetic phase separated condensates within bacterial cells.	50
Figure 18 - In vivo granules present similar dynamics as in vitro.	52
Figure 19 - QCP phase separation behavior in the absence or presence of slncRNAs	54
Figure 20 - Fluorescence recovery after photobleaching reveals different regimes.....	58
Figure 21 - RNP granule solidity can affect protein titer.	60
Figure 22 - slncRNA based protein expression leads to significant boost in titer	62

List of Tables

Table 1 - in vitro expression slncRNA sequences.....	14
Table 2 - in vivo expression slncRNA sequences.....	15

Abstract

Cellular compartments are essential for the organization of biological matter. Traditional compartments, such as the nucleus, are characterized by lipid membranes that separate them from their surrounding environment. In addition to these conventional organelles, there exists another category of compartments known as membraneless organelles which includes compartments such as paraspeckles and stress granules. Membraneless organelles, or bio-condensates in general, maintain distinct identities and perform various functions despite lacking a membrane. Recent research suggests that they self-assemble through the process of phase separation. In this process a homogenous solution segregates into multiple distinct phases, each with unique compositions and properties. Phase separation processes are classified by the material states (e.g., liquid, solid) of the different phases which can lead to various transitions (e.g., a liquid-solid transition indicates a liquid environment and a solid-like compartment).

Bio-condensates have been linked to numerous cellular activities, encompassing processes such as transcription, translation, and viral infection. The recognition of their significance has spurred a surge in research aimed at characterizing natural organelles and constructing synthetic compartments. The ability to design bio-condensates has significant implications for synthetic biology, particularly in enhancing biosynthetic processes and serving as drug delivery systems.

In our research we developed an RNA based synthetic bio-condensate. We first demonstrated that synthetic long non-coding RNA (slncRNA) molecules, which fold into secondary structures containing hairpin-like motifs, phase separate and form granules in-vitro. Using fluorescent RNA nucleotides, we showed that these slncRNAs form isolated puncta. Furthermore, the addition of a protein component which selectively binds the RNA hairpins resulted in almost complete co-localization between the proteins and slncRNAs. Thus, this simple two-component system constitutes a platform for bio-condensate formation.

We then developed a method to quantitatively measure the dynamics of solid-like granules by tracking their fluorescence signals over extended durations. We demonstrated both on simulated signals, and controlled experiments, that our algorithm successfully extracts important features from fluorescence data. Utilizing this algorithm, we investigated our bio-condensates and found that their properties depend on the number of hairpin motifs present in the RNA. Using disruption assays, we discovered that the granules protect the molecules inside from external influences. These findings suggest that our slncRNA-protein granules can selectively store proteins within them. Finally, we functionalized our bio-condensates by incorporating an additional component. Introducing a protein with a disordered region which enables multivalent interactions allowed the proteins to phase separate independently. We investigated the behavior of such condensates and found that this platform can be utilized to drastically overexpress proteins in bacteria.

Abbreviations and Notations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
RBP	RNA binding proteins / Receptor binding protein (for chapter 2 only)
CP	Coat protein
PCP	PP7 coat protein
QCP	Q β coat protein
MCP	MS2 coat protein
bp	Base pair
C4-HSL	N-butyryl-L-Homoserine lactone
LLPS	Liquid liquid phase separation
slncRNA	Synthetic long noncoding RNA
PBS	Phosphate buffered saline
AU	Arbitrary units
FP	Fluorescent protein
AMP	Ampicillin
KAN	Kanamycin
μ l, ml	Microliter, Milliliter
nM,mM,M	Nanomolar, Millimolar, Molar
OD	Optical density
PCR	Polymerase chain reaction
RT	Room temperature
FACS	Fluorescence assisted cell sorting
FRAP	Fluorescence recovery after photobleaching
FCS	Fluorescence correlation spectroscopy
MSE	Mean squared error
QQ plot	Quantile-Quantile plot
SE	Standard error
IDP/IDR	Intrinsically disordered protein/region

Introduction

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 milieu. In addition to these conventional membrane-bound organelles, there exists another category of cellular compartments known as membraneless organelles, and often referred to as biomolecular condensates. This category includes compartments such as paraspeckles, stress granules and Cajal bodies among others^{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 biomolecular condensates, often characterized as supramolecular assemblies of proteins and RNA molecules, self-assemble through the process of phase separation³⁻⁵.

Phase separation, or in the broader sense, phase transition, is a 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.

The increasing recognition of the significance of phase separated compartments has spurred a surge in research endeavors aimed at probing and characterizing natural biomolecular condensates, as well as constructing entirely synthetic compartments⁶⁻⁸. A primary method of investigation is *in vitro* reconstitution. This method entails incubation of individual components of condensates in specific conditions to test whether they phase separate *in vitro*^{6,9,10}. Alternative methods attempt to probe the dynamics of the already-formed condensate 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^{7,11,12}. 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.

Biomolecular condensates have been linked to numerous cellular activities across all forms of life, encompassing processes such as transcription, translation, viral infection, and extending to pathological conditions like neurodegenerative disorders^{1,13-16}. Of the many types of biomolecular condensates, ribonucleoprotein (RNP) granules have attracted significant interest from the scientific community¹⁷⁻²⁰, primarily due to their involvement in gene expression²¹. Furthermore, research has linked disruption in RNP granule function to the onset of specific cancers and neurological conditions¹⁵. RNP granules are widespread in eukaryotic cells, with notable examples including the nucleolus, stress granules and P-bodies. Granules were also recently found to exist in bacterial cells, where they play a role in mRNA degradation²².

RNP granules are broadly characterized as macromolecular assemblies of RNAs and proteins, with distinct granules containing different sets of macromolecules. The phase transition process of these granules is facilitated through multivalent RNA-RNA, RNA-protein, or protein-protein interactions. Notably, RNA molecules play an active role in both the formation and dissolution of granules. This has been demonstrated *in vitro*, where RNA in small concentrations was shown to promote granule formation, while RNA in concentrations higher than a certain threshold led to granule dissolution^{23,24}. In addition, it has been demonstrated that RNA can modify certain physical characteristics of granules, such as their three-dimensional structure²⁵⁻²⁸.

While RNAs play a crucial role in the formation and regulation of RNP granules, they represent only one aspect of the system as these granules cannot exist without a protein component. Key properties of granule-associated proteins include the existence of RNA-binding domains, and intrinsically disordered regions (IDRs). IDRs are protein segments which exhibit rapid conformational changes at the nano- to micro-second timescales, rendering them unlikely or incapable of forming a stable structure²⁹⁻³². Due to their unstructured nature, IDRs possess remarkable flexibility in their ability to form protein-protein interactions, enabling them to associate with various cellular partners^{33,34}. In the context of phase separation, IDRs have been consistently reported as being able to change the material properties of condensates, from solid-like to liquid-like, or vice versa³⁵⁻³⁸.

In 2017, Jain & Vale³⁹ reported on the phase transition process of RNA molecules associated with repeat expansion diseases, both *in vivo* and *in vitro*. These RNA sequences, comprised of dozens of triplet-repeats of CAG or CUG nucleobases, form intramolecular hairpin structures⁴⁰, which facilitate multivalent intermolecular interactions. The RNA granules presented features associated with liquid-solid phase transition systems: a lack of internal mobility, virtually no fusion events, and dependence on the number of repeats in the RNA sequence (i.e., cross linkers) rather than the concentration of the RNA. These characteristics helped establish the granules as physical solids.

Hairpin forming RNA sequences are widespread in the RNA world and are not strictly associated with disease phenotypes. Such sequences are commonly used in synthetic systems for biological research. Perhaps the most ubiquitous system is composed of RNA sequences that encode multiple hairpin motifs that can bind the phage coat proteins (CPs) of PP7 or MS2. Using this system to label the 5' or 3' end of a transcript has become commonplace in the last two decades⁴¹⁻⁴⁶, and enables visualization of RNA transcripts when the CPs are co-expressed. This approach, originally introduced by Singer and others⁴¹⁻⁴³, was devised for the purpose of probing the dynamics of transcription and other RNA-related processes, irrespective of cell-type. When co-expressed, the coat-protein-bound RNA molecules yield bright puncta, which are similar in appearance to natural biomolecular condensates.

In this research, we rely on our previous work⁴⁷ to design and synthesize a variety of PP7 coat-protein (PCP) binding, synthetic long non-coding RNA molecules (slncRNAs). Using fluorescent RNA nucleotides, we demonstrate that these slncRNAs form isolated puncta in vitro in a manner dependent on the number of hairpins encoded into the RNA. We further show that addition of fluorescent PCP to the suspension results in almost complete co-localization between protein and slncRNA.

We then developed a method to quantitatively measure the dynamics of phase separated gel-like granules by tracking their fluorescence signals over extended durations. 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.

Utilizing this algorithm on our slncRNA-based granules, we discovered that the various puncta emitted similar signals characterized by bursts of increasing or decreasing fluorescent intensity. We show that signal intensities and temporal characteristics are dependent on the number of hairpins present in the RNA. Using these observations, we conclude that these “fluorescence-bursts” corresponded to addition or shedding of slncRNA-PCP nucleoprotein complexes. These events occur at rates that are consistent with the puncta being phase-separated solid-like granules. Consequently, we present these slncRNA-protein granules as a genetically encoded platform for the selective storage of proteins.

Finally, we introduce a protein with an intrinsically disordered region to our system to increase its complexity. We hypothesized that this could lead to a different set of behaviors and provide another mechanism of control over granule characteristics and dynamics. To accomplish this, we introduced a coat protein with a well-documented disordered region, originating from the bacteriophage Q β . This IDR enables multivalent interactions between the proteins themselves. We demonstrate that the Q β coat protein can phase separate on its own when overexpressed in cells, and that introduction of hairpin containing RNA effectively modulates this behavior, yielding a complex RNA-valency dependent phase behavior. We explore the behavior of the condensates as a function of RNA valency as determined

by the number of hairpins encoded into the RNA and demonstrate that this platform can be utilized to drastically overexpress proteins in bacteria.

Background

Phase separation

From a theoretical perspective, phase separation can best be understood by examining the free energy of a solution, and its derivative, the chemical potential. These two properties are determined by factors such as molecular concentration, spatial distribution and the energy associated with chemical bonds between molecules. In the simplest scenario of non-interacting solute molecules in a solvent, the free energy exhibits a unimodal behavior, and the chemical potential demonstrates monotonic behavior as a function of solute concentration (Figure 1 – red curves). Under such conditions, the solute molecules are, on average, distributed homogeneously to maximize the entropy of the system. Thermal fluctuations that produce local inhomogeneities in concentration (and chemical potential) are counteracted by diffusive flux which equalizes differences in chemical potential across the system and minimizes free energy.

When solute molecules interact with one another, the free energy curve becomes multimodal, and the chemical potential behavior becomes non-monotonic, meaning some values of the chemical potential correspond to two different solute concentrations. In such a case, the system's free energy can be minimized by separating the molecules into two compartments with different concentrations but equal chemical potentials⁴⁸ (Figure 1 – blue curves).

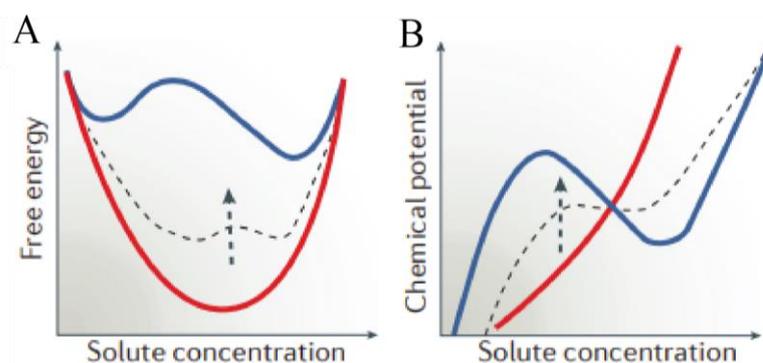


Figure 1 – Theoretical free energy and chemical potential curves.

A. Theoretical free energy curves and **B.** Theoretical chemical potential curves of different types of systems. Red curves– system with non-interacting molecules; Dashed black curves – one phase system with interacting molecules; Blue curves – Phase separated system with two compartments.

Figure from reference 48.

In molecular terms, all molecules exhibit varying degrees of weak, nonspecific interactions with each other and with the solvent (typically water in biological context). These interactions are characterized by low affinity, transient nature, and relatively equal distribution across the surface of the molecule. Given this, the concentration at which phase separation occurs is governed by a balance between weak interactions between the molecules themselves, and those between molecules and the solvent. When the interactions between the molecules are weaker than those with the solvent, the molecules will remain dispersed in solution at all concentrations. However, when the molecule-molecule interactions are stronger than molecule-solvent interactions, the system potential of the system to phase separate increases.

In such systems, phase separation occurs when the energetics favoring molecule-molecule interactions surpass the tendency of the solution to remain homogeneous. At this solubility limit, the molecular mixture separates into two phases, a dilute phase with a large volume and low concentration, and a condensed phase with small volume and high concentration. The phase separated state minimizes the free energy of the system and equilibrates the chemical potential between both existing phases. This essentially eliminates the net diffusive motion between the different phases, while allowing individual molecules to move between them⁴⁸.

Gelation

Gelation is a type of phase transition commonly encountered in biological systems. This transition refers to a phenomenon in which a solution transforms from a dispersed state of monomers or oligomers, into a cohesive three-dimensional network structure, termed a gel. This transition is driven by the physical, non-covalent cross-linking of specific associative domains, also referred to as sticker motifs. Gelation is governed by a critical concentration threshold, known as the percolation threshold which defines the point at which the concentration of associative domains reaches a critical value⁴⁹.

In biological context, the interactions mediated by the sticker motifs can be highly varied, these include polar interactions involving charges or dipoles, hydrogen mediated interactions, etc. The extent of cross-linking influences key properties of the gel, including its free volume, porosity, and average stiffness. Additionally, the timescales associated with the formation and disruption of these physical cross-links play a crucial role in determining the dynamic behavior of the gel⁵⁰⁻⁵².

Importance of phase transitions in cells

In cellular context, phase separation serves as the fundamental mechanism driving the formation of membraneless organelles. These organelles, characterized by their dynamic nature and absence of a surrounding lipid bilayer, have garnered increased attention in recent years owing to their pivotal involvement in diverse cellular processes^{1,2}. Eukaryotic cells harbor a myriad of organelles, each distinguished by a unique composition, spatial localization, and function⁵³ (Figure 2). Certain organelles are constitutively present in most cell types and contribute to essential housekeeping functions such as

ribosome biogenesis and RNA processing, while others assemble in response to specific environmental conditions⁵⁴. Membraneless organelles were also recently discovered in prokaryotic cells, where they contribute to pivotal processes such as RNA degradation and DNA organization⁵⁵⁻⁵⁸.

Owing to their critical roles, dysregulation of either the formation or functions of membraneless organelles is often associated with various pathological conditions, including neurodegenerative disorders and cancer⁵³. This highlights the importance of understanding the biophysical principles governing their formation and regulation.

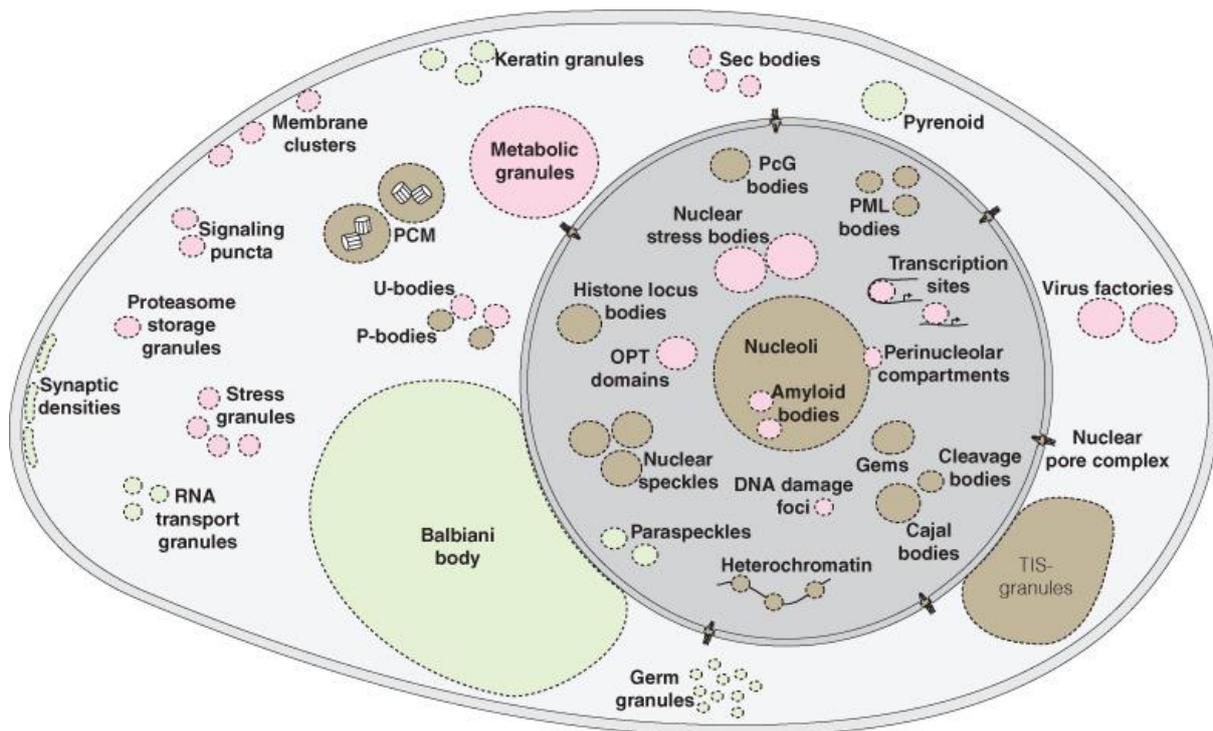


Figure 2 - Membraneless organelles in the cell.

Illustration of the known membraneless organelles in the Eukaryotic cell.

Figure from reference 53.

Polyphasic linkage theory

Given the complex nature of membraneless organelles and their various components, it is necessary to establish a theoretical framework to study the contribution of each molecular species. Polyphasic linkage theory, formulated by Wyman and Gill⁵⁹, has been extensively developed and utilized to explain how binding and linkage relations control phase transitions^{60,61}. This formalism considers multivalent molecules driving phase transitions as scaffolds, while molecules binding the scaffolds and controlling phase separation are known as ligands. Ligands do not undergo phase separation on their own and are not required for the phase separation of scaffolds.

Ligands that preferentially bind to the scaffold in the dilute phase suppress phase separation either by increasing the excluded volume of the scaffold, or by binding to sites that participate in scaffold-scaffold interactions, thereby reducing the scaffold's tendency to phase separate. Conversely, ligands that preferentially bind to the scaffold in the dense phase can enhance phase separation by increasing the partial concentration of the dense phase, or by serving as anchor points that can bind to additional scaffold molecules in the case of multivalent ligands (Figure 3 - Schematic of polyphasic linkage).

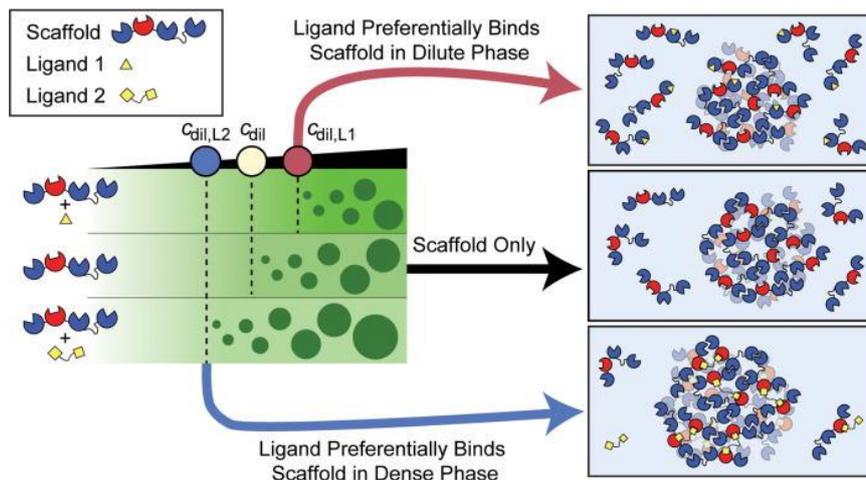


Figure 3 - Schematic of polyphasic linkage

According to polyphasic linkage theory, ligands can destabilize or promote phase separation according to their binding preferences to the scaffold molecule. This mechanism is used by the cell to regulate the phase separation process.

Figure from reference 60.

Intrinsically disordered protein regions

In the past, it was commonly accepted that the structure of a protein was directly related to its function, a principle known as the structure-to-function paradigm. This principle asserts that a protein must fold to a unique and stable 3D structure to carry out its biological activities. However, a considerable body of evidence has accumulated over the past few decades, challenging the validity of this paradigm. Although many proteins do conform to the paradigm and fold into a specific structure, a significant portion of the proteome comprises regions that are unlikely or unable to adopt a stable structure, but still exhibit functional activity. These regions are known as intrinsically disordered regions (IDRs), and proteins that are entirely disordered are referred to as intrinsically disordered proteins (IDPs).

The predominant theoretical model regarding the behavior of IDRs suggests that these regions exhibit dynamic conformational changes at the nano- to micro- second time scales. Due to their unstructured nature, IDRs possess remarkable flexibility in their protein-protein interactions, enabling them to form weak associations with various cellular partners. As a result, IDRs have been identified as critical

components in protein-protein interaction networks⁶² and naturally occurring liquid-liquid phase separated (LLPS) compartments⁶³.

Common characteristics of IDRs are well suited for driving phase transition processes. The lack of a stable secondary and tertiary structures allows them to engage in a wide range of interactions. In addition, IDRs frequently contain post translational modification sites, providing a mechanism for the cell to regulate phase behavior. As a result, biomolecular condensates often contain many proteins with IDRs⁶⁴. Interestingly, different species of condensates seem to be enriched with IDRs exhibiting different sequence features. For example, nucleoli contain arginine-rich IDRs, while nuclear speckles contain serine-rich IDRs. The different sequence features in different types of condensates suggest that IDRs can drive phase separation in more than one way⁶⁵.

RNA phase transitions in repeat expansion disorders

Trinucleotide repeat disorders comprise over 30 genetic disorders resulting from trinucleotide repeat expansion, a mutation characterized by an increase in copy numbers of three-nucleotide repeats (i.e., ‘CAG’, ‘CGG’, ‘CTG’, etc.). Well known examples of such disorders are Huntington’s disease and Friedreich’s Ataxia. This genetic mutation leads to developmental, neurological, or neuromuscular disorders. Generally, the magnitude of the repeat expansion correlates with the speed of disease onset and the severity of the resulting disorder^{40,66}.

Three hypotheses have been proposed to explain the mechanisms through which repeat expansions lead to pathogenicity. Firstly, repeat expansions might induce a gain or loss of function in the associated protein, assuming they occur in a coding region. However, it has been observed that mutations or deletion of the gene linked to the repeats do not consistently manifest the disease phenotype, whereas the expression of expanded repeats alone is sufficient to trigger a pathological state in model systems⁶⁷⁻⁶⁹. Secondly, RNAs containing repeats have the potential for non-canonical translation, producing short unnatural peptides which may lead to pathogenicity⁷⁰. Thirdly, repeat containing RNAs can accumulate in the nuclei, potentially sequestering various RNA binding proteins and disrupting cellular function.

In 2017, Jain and Vale³⁹ observed the formation of RNA granules in both in vivo and in vitro, from RNA sequences characterized by triplet repeats (e.g., CAG, CUG). Such RNA molecules are commonly associated with repeat expansion disorders. These sequences form intramolecular hairpin structures, which promote multivalent intermolecular interactions. The RNA granules displayed several properties indicative of liquid-solid phase transition systems, such as a lack of internal mobility, rare fusion events, and a dependency on the number of repeats rather than RNA concentration.

RNA stem-loops

RNA stem-loops, also referred to as RNA hairpins, are structural motifs commonly found in RNA molecules, characterized by a hairpin-like structure formed through base-pairing interactions within a single-stranded region of the RNA sequence. The stem of the structure consists of complementary base

pairs, forming a double-stranded helical region, while the loop represents an unpaired region connecting the base-paired stems⁷¹. RNA stem-loops are prevalent in diverse RNA molecules, including messenger RNAs, transfer RNAs, and ribosomal RNAs, as well as non-coding RNAs. In terms of functionality, stem-loops participate in many regulatory functions, such as binding to proteins, facilitating RNA-protein interactions, and serving as recognition sites for various cellular processes, including RNA splicing, translation, and RNA degradation⁷².

RNA stem-loops are also commonly used in synthetic systems for biological research. Perhaps the most widely used application is the MS2-tagging method which utilizes RNA molecules that encode for multiple hairpin motifs that can bind the phage coat proteins (CPs) of PP7 or MS2. Using this system to label the 5' or 3' end of a transcript enables visualization of RNA transcripts when the CPs are co-expressed (Figure 4). This approach, originally introduced by Singer and others, was devised for the purpose of probing the dynamics of transcription and other RNA-related processes, irrespective of cell-type⁴¹⁻⁴⁶.

One caveat of this tagging method is that many copies of the stem-loop motif need to be added to the untranslated region (UTR) of the target RNA to produce enough signal to view and track one RNA molecule in the nucleus. This raises a significant difficulty as DNA tandem repeats are notoriously hypermutable, often leading to destruction of the stem-loop binding site and reducing labelling efficacy and leading to imaging artifacts^{42,73}. For this reason, Katz et. al. developed a deep-learning based algorithm to generate synthetic stem-loop motifs with dissimilar sequences and high binding affinity to their associated coat protein⁴⁷. The synthetic stem-loops allowed construction of large 'cassettes' of 30 or more stem-loops that proved to be stable over time in cells.

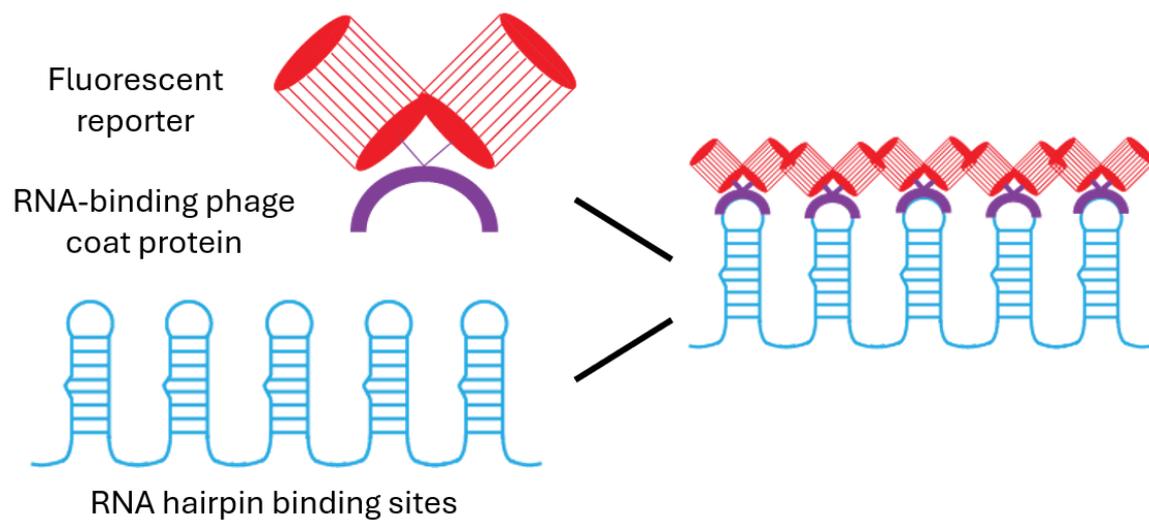


Figure 4 - RNA tagging using phage coat proteins.

Schematic of the “MS2 tagging” method. Expression of a fluorescent reporter fused to an RNA-binding phage coat protein, together with an RNA sequence which folds to a secondary structure containing multiple hairpin binding sites leads to tagging of the RNA upon binding of the coat proteins.

Phage coat proteins

The coat (capsid) proteins of single-stranded RNA bacteriophages such as MS2, PP7 and Q β , serve dual functions: they form the basis of the phage’s outer shell, encapsulating its genome, and act as translational repressors. The repression occurs by the selective binding to an RNA stem-loop structure in the phage genome, which includes the initiation codon of the phage replicase gene. By binding this structure, the coat proteins are thought to stabilize the RNA’s secondary structure, thereby inhibiting translation. Despite substantial differences in the sequences of the proteins and of the RNA loops which they bind, this dual functionality is conserved among these bacteriophages.

The selective binding property of these coat proteins has been extensively utilized for tracking and visualizing RNA molecules. This is done by cloning an RNA sequence encoding for multiple hairpins to the 5’ or 3’ of the target RNA and expressing the coat protein fused to a fluorescent reporter^{41,42}. Given that these proteins are known to interact with copies of themselves to form the complete phage capsid, modified versions have been developed for these applications, including in this work. For the PP7 coat protein, deletion of residues 67-75 (known as the FG region), effectively nullifies the ability of the protein to form the full viral capsid⁷⁴. For the Q β coat protein, it was found that fusion of a peptide longer than 20 residues to the coat protein’s C-terminus, prevents capsid formation⁷⁵⁻⁷⁸. Consequently, all versions of the Q β coat protein used in this work are fused to an mCherry fluorescent protein at their C-terminus.

Research Objectives

Goals

1. Investigate the phase separation potential of RNA molecules containing hairpin like motifs.
2. Design a method to study the dynamics and signal characteristics of non-liquid-like phase separated compartments.
3. Characterize the behavior of RNA hairpin-based condensates as a function of the number of hairpins present on the RNA, and their architecture.
4. Investigate the behavior of RNA-protein condensates containing proteins with an intrinsically disordered region.

Significance

Phase separation is now recognized as a crucial mechanism by which cells spatially and temporally organize components and reactions. This understanding has spurred significant efforts to create synthetic phase-separated condensates for various functions. These efforts typically involve the use of well-characterized proteins that drive phase separation, combined with additional components. Engineering completely synthetic RNA-based condensates that can form independently of protein components and selectively incorporate proteins would expand the current toolkit of synthetic condensates. This advancement would enable the development of condensates with programmable characteristics for a wide range of applications.

Materials and Methods

Bacterial strains

E. coli BL21-DE3 cells which encode the gene for T7 RNAP downstream from an inducible pLac/Ara promoter were used for all reported experiments. *E. coli* TOP10 (Invitrogen, Life Technologies, Cergy-Pontoise) was used for cloning procedures.

Addgene plasmids

pCR4-24XPP7SL was a gift from Robert Singer (Addgene plasmid # 31864; <http://n2t.net/addgene:31864>; RRID: Addgene_31864).

pBAC-lacZ was a gift from Keith Joung (Addgene plasmid # 13422; <http://n2t.net/addgene:13422>; RRID: Addgene_13422).

Construction of slncRNA plasmids for *in vitro* expression

All sequences encoding for the *in vitro* slncRNAs (i.e., PP7-3x, PP7-4x, PP7-3x/MS2-3x, PP7-4x/MS2-4x, PP7-8x and PP7-14x/MS2-15x, Table 1) were ordered from Integrated DNA Technologies (IDT)

(Coralville, Iowa) as gBlock gene fragments downstream to a T7 promoter and flanked by EcoRI restriction sites on both sides. The gBlocks were cloned into a high copy pCMV plasmid containing an Ampicillin resistance gene and verified using Sanger sequencing.

Table 1 - *in vitro* expression slncRNA sequences

slncRNA	DNA sequence
PP7-3x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> cacggaggatgcgggaaacatgaag atcactcatgttcgcttaacatggatagggatcacccatgttcggtggtgcgtcaacCAGAGAT <u>TTCATATGGGAAACTCTGggacacgctgtattatacatgaggatcacccatgtgtgcttaaat</u> <u>atgggtaATTCCAGTTTATATGGAAACGGA</u>
PP7-4x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> cacggaggatgcgggaaacatgaag atcactcatgttcgcttaacatggatagggatcacccatgttcggtggtgcgtcaacCAGAGAT <u>TTCATATGGGAAACTCTGggacacgctgtattatacatgaggatcacccatgtgtgcttaaat</u> <u>atgggtaATTCCAGTTTATATGGAAACGGAAT</u> aagattagatatggtttAAACGACA <u>ATATATGGATTGCGTTT</u> tggggcacgctctggttactgcgatgggtgggaggttagaggga ttctaagcttaaga
PP7-3x/MS2-3x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> gtggaacgtaatggacaTGAAGA CGATTACGCTTCacacggaggatgcgggaaacatgaagatcacccacgcttaacatggata gggatcacccatgttcggtggtgcgtcaacCAGAGATTTTCATATGGGAAACTCTGggac acgctgtattatacatgaggatcacccatgtgtgcttaaatatgggtaAGTTGACCATTAGGCA ACTgtaagatgctccggttaATTCCAGTTTATATGGAAACGGAATtgatgtaccgttgag caAGAACACGATTACGGTTCtcgattagatatggggtgggaggttagagggatctcgca gaaga
PP7-4x/MS2-4x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> gtggaacgtaatggacaTGAAGA CGATTACGCTTCacacggaggatgcgggaaacatgaagatcacccatgttcgcttaacatgg atagggatcacccatgttcggtggtgcgtcaacCAGAGATTTTCATATGGGAAACTCTGg gacacgctgtattatacatgaggatcacccatgtgtgcttaaatatgggtaAGTTGACCATTAG GCAACTgtaagatgctccggttaATTCCAGTTTATATGGAAACGGAATtgatgtaccgt tgagcaAGAACACGATTACGGTTCtcgattagatatggtttAAACGACAATATATGG ATTGCGTTTtggggcacgctctggAGAAGACCATTAGGCTTCTtactgcgatgggtg ggaggttagagggatctcgcaagaaga
PP7-8x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> cacggaggatgcgggaaacatgaag atcacccatgttcgcttaacatggatagggatcacccatgttcggtggtgcgtcaacCAGAGAT <u>TTCATATGGGAAACTCTGggacacgctgtattatacatgaggatcacccatgtgtgcttaaat</u> <u>atgggtaATTCCAGTTTATATGGAAACGGAAT</u> taggctagagcatggcggtattgagttc gggttgAAACGACAATATATGGATTGCGTTTtggggcacgctctggAAAGGAGA TTATATGAAATCCCTTgcgcgcaaccgggtagaTACGAGTCAATATGGTGACCGT AccttggtggggaagcgaTAAGCACATTATAAGGAATGGCTTAgtgctagaccctggtct TTTCGAGAAAATATGGTTCCGAAAatgggtgggaggttagaggattctcgcaagaaga a
PP7-14x/MS2-15x	<u>gagaaacgTTTCGACATTATATGGAATGCGAAA</u> gtggaacgtaatggacaTGAAGA CGATTACGCTTCacacggaggatgcgggaaacatgaagatcacccatgttcgcttaacatgg atagggatcacccatgttcggtggtgcgtcaacCAGAGATTTTCATATGGGAAACTCTGg gacacgctgtattatacatgaggatcacccatgtgtgcttaaatatgggtaAGTTGACCATTAG GCAACTgtaagatgctccggttaATTCCAGTTTATATGGAAACGGAATtgatgtaccgt tgagcaAGAACACGATTACGGTTCtcgattagatatggtttAAACGACAATATATGG <u>ATTGCGTTTtggggcacgctctggAGAAGACCATTAGGCTTCT</u> tactgcgaccgca ataAAAGGAGATTATATGAAATCCCTTgcgcgcaaccgggtagaAGATCACCAATT AGGGATCTgtaactcacggcgtatTACGAGTCAATATGGTGACCGTaaagctaggg catgtgccAGAAGAGCATTAGCCTTCTccttggtggggaagcgaTAAGCACATTATAA <u>GGAATGGCTTA</u> aagtggtgcggcggggACTTGACCATTAGGCAAGTgtgtagacc ctggtctTTTCGAGAAAATATGGTTCCGAAAgaactatacgaagtgcATGCGAGG ATTACCCGCATatggtgcaaatgggagaATTGGAGTAAATATGGTTACCCAATagg ctagagcatgacggCAGTGAGAATTATCCACTGgttagcgggttaccgagATTGCACAT

	<p>TATATGGAATGGCAATgattcatgccggtcgtTTGTGAGGAGTACCCACAAaatgg gagggtgctataTAACCAGGTTATATGCAACCGTTAggcccgtttagtttAGTTC AGCATTAGCGAACTgtgcaagaccggtggcTAAGGAGTTTATATGGAAACCCCTT AgccctcgagcatgctaACATGAGGATTACCCATGTgatgggttgaaacgtgCAATTAT TTGATATGGCAAAAATTGgtagggcatggctgcaGCGTGAGAATTATCCACGcta ggctagagcatggcggtattgagttcgggtttgAGATGACCATTAGGCATCTgtgctagagc atgcaaaacgacataatatggtatgcttttggggtgctagtagtataaccACATGATGAGCACAC ATGTatgggtgggaggtagaggattctcgcgagaaga</p>
--	---

Binding site sequences are marked in capital letters: Underlined: PP7; Italicized: MS2

Construction of slncRNA plasmids for *in vivo* expression

The PP7-4x/Q β -5x slncRNA sequence (Table 2) was ordered from GenScript, Inc. (Piscataway, NJ), as part of a puc57 plasmid, flanked by EcoRI and HindIII restriction sites. pBAC-lacZ backbone plasmid was obtained from Addgene (plasmid #13422). Both insert and vector were digested using EcoRI-HF (NEB, #R3101S) and HindIII (NEB, #R3104S) and ligated to form a circular plasmid. The sequence was verified by sanger sequencing.

The sequences for Q β -1x/3x/5x/10x/12x/17x/NS (Table 2) were generated using the CARBP online tool⁴⁷ and ordered IDT as gBlock gene fragments downstream to a T7 promoter and flanked by ApaI and AvrII restriction sites. The gBlocks were cloned into a pPROLAR vector containing a kanamycin resistance gene, which was linearized using the above-mentioned restriction enzymes (NEB, #R0114S and #R0174S, respectively), and excised from 1% agarose gel following electrophoresis to obtain the relevant DNA segment. Sequences were verified using Sanger sequencing.

The DNA sequence of the TagBFP protein was ordered from IDT as a gBlock gene fragment flanked by AvrII and AgeI restriction sites and ligated to the slncRNA-containing pPROLAR vectors linearized using the same restriction enzymes (NEB, # R0174S and #R3552S, respectively).

Table 2 - *in vivo* expression slncRNA sequences

slncRNA	DNA sequence
PP7-4x/ Q β -5x	<p>gagaaacctaggcgattatgacgttattctactttgattgtgATGCATGTCTAAGACAGCATcgccctgctggt cgtgacTAAGGAGTTTATATGGAAACCCCTTAcgagacaatgctacctaccggtcgggccactgttttt accctatgATGCATGTCTAAGACAGCATcgccctgctggtcgtgacTAAGGAGTTTATATGGAAACC CTTAgaaacagccctgccttgaagccgagaacaATGCATGTCTAAGACAGCATatggattgcctgctctg tTAAGGAGTTTATATGGAAACCCCTTAcacagccttcgagtagtcaacgcttgcATGCATGTCTA AGACAGCATtcaccgcttctcaagTAAGGAGTTTATATGGAAACCCCTTAgctactaactcgagAT GCATGTCTAAGACAGCATcagaaactcacgtcctggctgtcga</p>
Q β -NS	<p>CAGCGCTTATTCCAAAAGAAGGTTAGAACgGCTTTAGGAAAGCAACACATTTTATGACAT TAGAAAGGGTCGTtTTGCCaGGAGCgGAGTTCTGTCTGAGGGAGATGGCAATGTATGCA AGTGGAGTACGTAAGGGTgATGCACTAAATAAGGTACGATCgGTGCATCTACCTGGGTC ATAaACaACtCGAACTTAAGAATGaTgGAGGGGGTAAAAGTAATTGTGGCCCaAGTTaTT AGCGggATCTATCCAtCCcAGAGCAGGCTATTGCTaGAGGGCTCCTGTACaTGTGTGAG CAAGTAtGTGGAgAAATTAGACgAGGGCTaCctGCagTAGgTAGACTGTGgTGACCAAGG TAGAGATGTCCTCTAATGCGGAGTGCCAGTCTaTCGACGCCGATTCTCGCGAgaa</p>

Qβ-1x	CAGCGCTT <u>aaggaaactgtaacagtcctt</u> ATTCCAAAAGAAGGTTAGAACgGCTTTAGGAAAGCA ACACATTTTATGACATTAGAAAGGGTCGTtTTGCGaGGAGCgGAGTTcGTCTGAGGGAG ATGGCAATGTATGCAAGTGGAGTACGTAAGGGTgATGCACTAAATAAGGTACGATCgGT GCATCTACCTGGGTCATAaACaActCGAACTTAAGAATGaTgGAGGGGGTAAAAgTAATT GTGGCCCaAGTTaTTAGCGggATCTATCCATtCCcAGAGCAGGCTATTGCTaGAGGGCTCC TGTACaTGTGTGAGCAAGTAtGTGGAgAAATTAGACgAGGGCTaCctGCagTAGgTAGACT GTGgTGACCAAGGTAGAGATGTCCTCTAATGCGGAGTGCCAGTCTaTCGACGCCGATTTC TCGCGAga
Qβ-3x	CAGCGCTT <u>aaggaaactgtaacagtcctt</u> ctcgtgctgatcgtgacttggatgtccaagacaccaaCGAGACA ATGCTACCTTACCGTCGGCCCACTTGTTTTACCCATG <u>acatgacgagatactcgcatgt</u> ATTCCA AAAGAAGGTTAGAACgGCTTTAGGAAAGCAACATTTTATGACATTAGAAAGGGTCGT tTTGCGaGGAGCgGAGTTcGTCTGAGGGAGATGGCAATGTATGCAAGTGGAGTACGTA AGGGTgATGCACTAAATAAGGTACGATCgGTGCATCTACCTGGGTCATAaACaActCGAA CTTAAGAATGaTgGAGGGGGTAAAAgTAATTGTGGCCCaAGTTaTTAGCGggATCTATCC ATtCCcAGAGCAGGCTATTGCTaGAGGGCTCCTGTACaTGTGTGAGCAAGTAtGTGGAgA AATTAGACgAGGGCTaCctGCagTAGgTAGACTGTGgTGACCAAGGTAGAGATGTCCTCT AATGCGGAGTGCCAGTCTaTCGACGCCGATTCTCGCGAga
Qβ-5x	gagaaa <u>aaggaaactgtaacagtcctt</u> ctcgtgctgatcgtgacttggatgtccaagacaccaaCGAGACAATG CTACCTTACCGTCGGCCCACTTGTTTTACCCATG <u>acatgacgagatactcgcatgt</u> tcgctgctggtc gtgac <u>atgcatgtctaagacagcat</u> GAAACAGCCGTCGCCTTGAAGCCGAGAACA <u>tgcatgtcgaag</u> <u>acagcaatggaa</u>
Qβ-10x	GAGAAA <u>aaggaaactgtaacagtcctt</u> CTCGTGCTGATCGTGACTtggatgtccaagacaccaaCGAGA CAATGCTACCTTACCGTCGGCCCACTTGTTTTACCCATG <u>acatgacgagatactcgcatgt</u> TCGCC TGCTGGTCGTGAC <u>atgcatgtctaagacagcat</u> GAAACAGCCGTCGCCTTGAAGCCGAGAACA <u>t</u> <u>gcatgtcgaagacagcaa</u> ATGGATTGGGTCTCCAATTCCTGTCTGT <u>ttccatgactaagt</u> <u>caggaa</u> C ATCAGGCTTCGCAGTATGCAACGCTTGCg <u>atgattgcaaagcaagcat</u> TTCACCGCTTTCCTA AG <u>aaggatagtaatgactacctt</u> GTAATACTCGCAG <u>atcgaactctaagagtcgat</u> CAGAAAACGTCAC GTCCTGGC <u>aacatgtcagggacaggtt</u> TGGAA
Qβ-12x	GAGAAA <u>aaggaaactgtaacagtcctt</u> GAGCTTAACCATGCATG <u>aatcgagaaaat</u> <u>atggttccgatt</u> G GCCGTTGTGTTAGTT <u>gaataaggattacctattc</u> ATGGATTGCCTGTCTGT <u>taagacagattactgc</u> <u>tta</u> GAGACACGCTGATTTATACATGAGGATCACCATGTGTGCTTAAATATGGGTA <u>taaggac</u> <u>tttatatgtaaagcctta</u> TCGATTAGATATGGTT <u>acataaggattacctatgt</u> GCTAGAGCATGAG GAC <u>cgtaataattatatacgg</u> GGCGATGGCTGACGTAG <u>atacagttctaagaacgat</u> TGGCTAGAGC ATGTGATACATGAGGTCACCCATGTGGTATTTGATCGTGGGCa <u>atgcacatgtaaacatggcatt</u> GTGCTAGAGCATGCGA <u>aatgcacattatggaatggcatt</u> AAGTGGTGC GGCGGGGGT <u>cagatt</u> <u>tcatatgggaaactgta</u> GGGCGTCGAGCGAATATACATGAAGATCACCCATGTGGGCTAGAG CATGGTAT <u>taaggagt</u> <u>ttttatgtaaacctta</u> GTGGAA
Qβ-17x	GAGAAA <u>aaggaaactgtaacagtcctt</u> TTACTGCGACCGCAATA <u>aatgcacattat</u> <u>atggaatggcatt</u> AC CTGTAACAGGGGGAG <u>ccgtaataattatatacgg</u> ATGGATTGCCTGTCTGT <u>tacagaacttatatgg</u> <u>aagtctgta</u> AAGTGGTGC GGCGGGGGT <u>taaggagttaagttaaacctta</u> AGGCTAGAGCATGAC GG <u>taaggagt</u> <u>ttttatgtaaacctta</u> GTTAGCGGGTTACCGAG <u>atgcatgtcaagacagcat</u> CGCCT GCTGGTCGTGAC <u>aaggatagtaatgactacctt</u> TGGCTAGAGCATGTGATACATGAGGTCACCC ATGTGGTATTTGATCGTGGGCa <u>aacatgtctgagacagttt</u> GGGCGTCGAGCGAATATACATG AAGATCACCCATGTGGGCTAGAGCATGGTAT <u>tacagatttcatatgggaaactgta</u> TTATTGTTA GAGTGAGC <u>atacagattatggaatccgat</u> GGGCTAGAGCATGCGT <u>agtgaatatctaagatatca</u> <u>c</u> GAGTGTCCGGGAAGTCT <u>atccatgtcaagacagcat</u> GGTAGGGCATGGCTGCA <u>atttacttcta</u> <u>agaagaat</u> TAGGCTAGAGCATGGCGG <u>agatatccattcggtatct</u> AAGCTAGGGCATGTGCC <u>gt</u> <u>acacgattacggtaca</u> ATGGGTGGGAGGTAGAG <u>taaggagt</u> <u>ttcacagtaaacctta</u> TGGAA

Binding site sequences are marked: Underlined: PP7; red colored: Qβ

Design and construction of fusion-RBP plasmids

RNA binding proteins (RBP) sequences lacking a stop codon were amplified via PCR off either Addgene or custom-ordered templates. Both RBPs presented (PCP and QCP) were cloned into the RBP plasmid between restriction sites KpnI and AgeI, immediately upstream of an mCherry gene lacking a start codon, under the RhIR promoter containing the rhIAB las box⁷⁹ and induced by N-butyryl-L-homoserine lactone (C4-HSL) (Cayman Chemicals, Ann Arbor, Michigan). The backbone contained either an Ampicillin (Amp) or Kanamycin (Kan) resistance gene, depending on the experiment.

In vitro transcription of slncRNA

A vector containing the slncRNA DNA sequence, flanked by two EcoRI restriction sites, was digested with EcoRI-HF (NEB, #R3101S) per the manufacturer's instructions to form a linear fragment encoding the slncRNA sequence. The enzyme was then heat-inactivated by incubating the restriction reaction at 65° C for 20 minutes. For fluorescently labelled RNA, 1 µg of the restriction product was used as template for in vitro transcription using HighYield T7 Atto488 RNA labeling kit (Jena Bioscience, Jena, Germany, RNT-101-488-S), according to the manufacturer's instructions. Non-fluorescent RNA was transcribed using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB, #E2040S). Following *in vitro* transcription by either kit, the reaction was diluted to 90 µl and was supplemented with 10 µl DNase I buffer and 2 µl DNase I enzyme (NEB #M0303S) and incubated for 15 minutes at 37° C to degrade the DNA template. RNA products were purified using Monarch RNA Cleanup Kit (NEB, #T2040S) and stored in -80°.

Protein expression and purification

A bacterial plasmid encoding his-tagged tdPP7-mCherry under the rhIR promoter (containing the las box, inducible by N-butyryl-L-Homoserine lactone [C4-HSL], Cayman Chemical), ampicillin resistance, and RhIR was transformed into E. coli TOP10 cells (Invitrogen Life Technologies, Cergy-Pontoise). Cells containing the plasmid were grown in 10 ml Luria-Bertani medium (LB: 10 g NaCl, 10 g tryptone, and 5 g yeast extract in 1 L deionized water, autoclaved) containing 100 µg/ml ampicillin (Amp) in a 50 ml falcon overnight, at 37 °C and 250 rpm. The culture was diluted into 500 ml 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) containing 100 µg/ml Amp and 97 nM C4-HSL in a 2-liter flask and grown for another day at 37 °C and 250 rpm. Culture was visibly pink the next morning. Cells were centrifuged at 8000 rpm for 10 min in 250 ml bottles, supernatant was discarded, and the visibly pink pellets were resuspended in resuspension buffer (RB: 50 mM Tris, 100 mM NaCl, 0.02% sodium azide in deionized water, pH 7.0). The resuspended cells were lysed by passing the culture four times through a high-pressure homogenizer (Emulsiflex, Avestin Inc, Canada) at an average working pressure of 10-15 KPSI and maintained at 4 °C using a circulating bath (GMBH, Germany). The collected lysate was centrifuged at 13 KRPM for 30 min. Clear, visibly pink supernatant was collected,

and cell debris was discarded. The supernatant was incubated with Ni-coated beads (HisLink protein purification resin, Promega) at room temperature for 1 hr, with 13 rpm overhead rotation. The his-tagged proteins were then purified on a gravity-flow column (PolyPrep chromatography column, Biorad) using the manufacturer's protocol. The buffer of the eluted protein was changed to phosphate buffered saline (PBS: Dulbecco's phosphate buffered saline -calcium -magnesium, Biological Industries) by rinsing multiple times with 1x PBS on a 3 kDa MWCO spin column (Amicon Ultra 0.5 mL, Merck Millipore).

In vitro granule preparation

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 hour. 3-5 μ l from the reaction was then deposited on a glass slide prior to microscopy.

For the RNase experiment, granules were first formed as described and allowed to rest at room temperature for 1 hour. Following this, RNase A enzyme (ThermoFisher, catalog: EN0531) was added at 35 nM final concentration. The reaction was then immediately deposited on a glass slide and proceeded to imaging.

For the competition experiment, granules were first formed with tdPCP-mCherry at a final concentration of 40 nM and allowed to rest at room temperature for 1 hour. Following this, tdPCP-mCerulean was added to the reaction at 80 nM final concentration prior to imaging.

Bacterial culture growth

BL21-DE3 cells expressing the two plasmid system (single copy plasmid containing the binding sites array, and a multicopy plasmid containing the fluorescent protein fused to an RNA binding protein) were grown overnight in 5 ml LB, at 37° with appropriate antibiotics (Cm, Amp), and in the presence of two inducers – 1.6 μ l Isopropyl β -D-1-thiogalactopyranoside (IPTG) (final concentration 1 mM), and 2.5 μ l C4-HSL (final concentration 60 μ M) 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 \times PBS buffer pH 7.4, 950 ml DDW) and 5% LB with appropriate antibiotics and induced with 1 μ l IPTG (final concentration 1 mM) and 1.5 μ l C4-HSL (final concentration 60 μ M). For stationary phase tests, cells were diluted into 3 ml Dulbecco's phosphate-buffered saline (PBS) (Biological Industries, Israel) with similar concentrations of inducers and antibiotics. Culture was shaken for 3 hours at 37° before being applied to a gel slide [3 ml PBSx1, mixed with 0.045g SeaPlaque low melting Agarose (Lonza, Switzerland), heated for 20 seconds and allowed to cool for 25 minutes]⁸⁰.

1.5 µl cell culture was deposited on a gel slide and allowed to settle for an additional 30 minutes before imaging.

In vitro granule microscopy

Granules were imaged in a Nikon Eclipse Ti-E epifluorescent microscope (Nikon, Japan) with a 100x 1.45 NA oil immersion objective. Excitation was performed by a CoolLED (Andover, UK) PE excitation system at 488 nm (Atto 488) for experiments containing fluorescent RNA, 585 nm for mCherry protein, and 405 nm for cerulean protein. Images were captured using the Andor iXon Ultra EMCCD camera with a 250 msec exposure time for 488 nm, 250 msec exposure time for 585 nm, and 2 seconds for 405 nm.

Live cell microscopy for granule dynamics measurements

Gel slide was kept at 37° inside an Okolab microscope incubator (Okolab, Italy). A time lapse experiment was carried out by tracking a field of view for 60 minutes 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 at 585 nm (mCherry) wavelength by a CoolLED (Andover, UK) PE excitation system.

Quantification of the fraction of cells presenting puncta was done by taking 10-15 snapshots of different fields of view (FOV) containing cells. The number of cells showing puncta and the total number of fluorescent cells in the FOV were counted manually.

Structured Illumination super resolution microscopy

Super resolution images were captured using the Elyra 7 eLS lattice SIM super resolution microscope (Zeiss, Germany) with an sCMOS camera, a 63x 1.46 NA water immersion objective, with a 1.6x further optical magnification. 405 nm, 488 nm and 585 nm lasers were used for excitation of the cerulean, Atto-488, and mCherry respectively. 16-bit 2D image sets were collected with 13 phases and analyzed using the SIM² image processing tool by Zeiss.

Fluorescence recovery after photobleaching (FRAP) measurements

For fluorescence recovery after photobleaching (FRAP) measurements, cells were grown overnight with appropriate antibiotics and diluted 1:100 in the morning in a solution of BA-LB (95%-5% v/v) with the relevant induction (0.33 mM IPTG and 60 µM C4-HSL) and allowed to grow for 3-4 hours prior to measurement.

Gel slides for bacterial setting were prepared as follows: A solution of 1.5% (w/v) low melting agarose (Lonza, Switzerland) in PBS was heated for approximately 20 seconds to allow full melting of the

agarose. Approximately 50 μ l was deposited on a microscope slide, and a coverslip was immediately applied. When the agarose had hardened, the coverslip was carefully removed, leaving a thin pad of agarose on the slide. Ten microliters of a bacterial culture were then deposited on a fresh coverslip, and the microscope slide and pad were placed on top with the agarose facing down so that the culture was spread between the pad and coverslip. Translucent nail polish was then applied to the edges of the coverslip to seal the liquid and prevent evaporation.

FRAP measurements were done using an LSM-880 laser scanning confocal microscope (Zeiss, Germany) with an sCMOS camera, and a 63x 1.46 NA oil immersion objective. Following identification of a live cell, photobleaching of a localized region was done using a 488 nm laser at full power (25 mW). mCherry fluorescence recovery was measured using excitation with a 594 nm laser with 1% power (0.1 mW) with images taken every 15 seconds for a total of 5 minutes. FRAP image analysis was done using the ImageJ software with the Stowers plugin suite.

Flow cytometry measurements

IPTG / C4HSL-induced or non-induced *E. coli* BL21 cells were grown overnight at 37^o with appropriate antibiotics. In the morning, the bacterial culture was diluted 1:100 into 3 ml of 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 \times PBS buffer pH 7.4, 950 ml DDW) and 5% LB supplemented with the relevant inducers (IPTG and/or C4-HSL). The diluted cells were shaken for several hours prior to measurement to allow sufficient time for induction and protein expression. Samples and appropriate controls were loaded onto a 96-wells plate (Thermo Scientific, cat. 167008) in triplicates, each well containing 200 μ l of diluted bacterial cells. The cells were then measured using flowcytometry (MACSquant VYB, Miltenyi Biotec), with the 561 nm excitation laser and the Y2 detector channel (a 615/20 nm filter). The flow cytometer was calibrated using MacsQuant calibration beads (Miltenyi Biotec) before measurement. Running buffer, washing solution, and storage solution were all purchased from the manufacturer (catalog numbers 130092747, 130092749, and 130092748, Miltenyi Biotec, respectively).

Events were defined using an FSC-height trigger of 3. An FSC-area over SSC-area gate (gate-1) was created around the densest population in the negative control (non-induced cells), and events falling inside this gate were considered live bacteria. From the gate-1-positive cells, an SSC-height over SSC-area gate (gate-2) was created along the main diagonal, and events falling inside this gate were considered single bacterium. Finally, from the gate-2-positive cells, a histogram of mCherry distribution was created, with the threshold for mCherry-positive cells set by leaving around 0.1% positive events in the negative control.

Image analysis for *in-vitro* and *in-vivo* granule microscopy

The brightest spots (top 10%) in the field of view were tracked over time and space via the imageJ MosaicSuite plugin⁸¹⁻⁸³. A typical field of view usually 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 fed to a custom built Matlab (The Mathworks, Natick, MA) script designed to normalize the relevant spot data. Normalization was carried out as follows: for each bright spot, a 14-pixel wide sub-frame was extracted from the field of view, with the spot at its center. Each pixel in the sub-frame was classified to one of three categories according to its intensity value. The brightest pixels were classified as 'spot region' and would usually appear in a cluster, corresponding to the spot itself. The dimmest pixels were classified as 'dark background', corresponding to an empty region in the field of view. Lastly, values in between were classified as 'cell background'. We note that for the *in vitro* experiments the 'dark background' and 'cell background' pixel groups yield similar intensity values. This, however, does not affect the performance of the algorithm for *in vitro* experiments. Classification was done automatically using Otsu's method⁸⁴. 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 value and cell intensity value. This was repeated for each spot from each frame in the data, resulting in sequences of intensity vs. time for the spot itself and for the cell background.

Results

Chapter 1 – Formation of synthetic RNA and RNA-protein granules

Hairpin containing RNA phase separates in vitro into gel-like granules

To test whether hairpin containing RNA can phase separate in vitro we designed six synthetic long non-coding RNA (slncRNA) binding-site cassettes using our binding site resource^{47,85,86}. We divided our slncRNAs into two groups. For the first group (class I slncRNAs), we designed three cassettes consisting of three, four, or eight hairpins that encode for PCP binding sites (PCP-3x, PCP-4x, and PCP-8x, respectively). In this group, hairpins were spaced by a randomized sequence that did not encode for a particular structure. For the second group (class II slncRNAs), we encoded three cassettes that consisted of three, four, and fourteen PCP binding sites that were each spaced by hairpin structures that do not bind PCP (PCP-3x/MCP-3x, PCP-4x/MCP-4x, and PCP-14x/MCP-15x, respectively). In addition, we designed a negative control slncRNA which does not contain any hairpin binding sites. To ensure that the negative control sequence has a similar GC content as the other slncRNA molecules (45%), it was designed as a permutation of the PCP-8x sequence. The sequences encoding for the slncRNAs were cloned downstream to a pT7 promoter and transcribed in vitro to generate the corresponding RNA. To visualize the RNA, we incorporated fluorescent nucleotides in the transcription reaction such that an estimated 35% of uracil bases were tagged by Atto-488 fluorescent dye. Each slncRNA-type was separately mixed with granule forming buffer (Figure 5A) at equal concentration (8.5 nM final concentration) and incubated for 1 hour at room temperature. 2-5 μ l of the granule reaction were then deposited on a glass slide and imaged using an epi-fluorescent microscope.

The images show that reactions with slncRNA molecules which contain hairpin binding sites result in the formation of a multitude of bright localized fluorescent condensates, with the exception of the PCP-3x case (Figure 5B). Interestingly, when increasing the slncRNA concentration in the PCP-3x case, sporadic granules do begin to form at 20 nM, and more robustly at 40 nM (Figure 5C), reminiscent of a concentration dependent, liquid-like phase separation rather than gel-like. In contrast to the above, the granule reaction containing the negative control slncRNA does not result in any discernible puncta (Figure 5D). In addition to the localized, small condensates, we note the formation of larger structures in the reactions prepared with the longer slncRNA molecules (e.g., PCP-8x and PCP-14x/MCP-15x), consistent with a gel like solid network (Figure 5E).

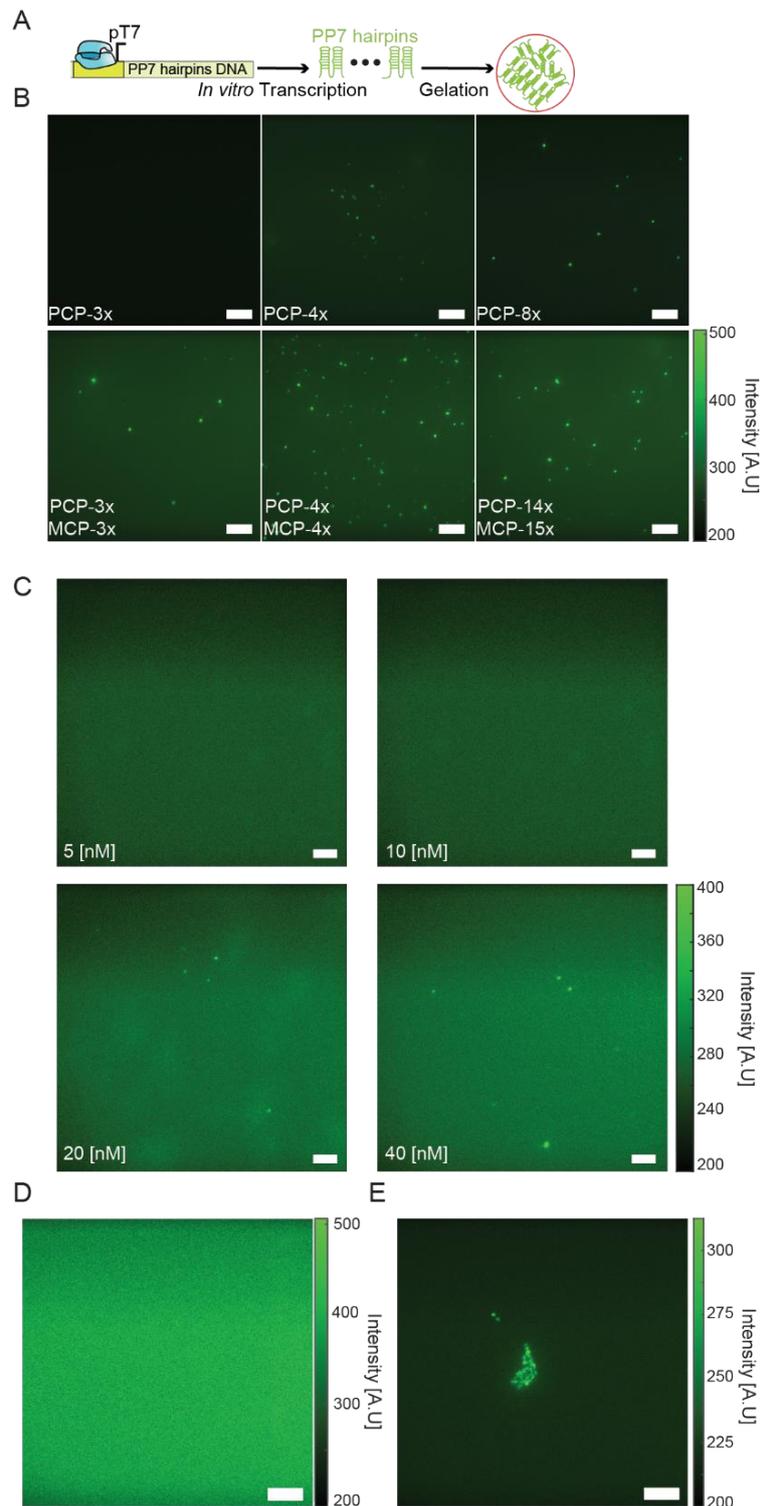


Figure 5 - In vitro RNA-only granules microscopy.

A. Construct diagram depicting in vitro transcription of hairpin containing slncRNA molecules used and their gelation under suitable conditions. **B.** Microscopy images showing dependence of structure morphology on the number of binding sites in the slncRNA. PP7-3x results in no visible puncta, while other slncRNAs shows multiple isolated puncta and additional larger fluorescent structures. All scale bars are 10 μ m. **C.** Sample microscopy images of granule reactions with different concentrations of PCP-3x slncRNA. The images show the formation of weak granules at concentration of 20 nM, and brighter, larger granules at 40 nM. Concentration written on the images. Scalebars are 10 μ m. **D.** Negative control slncRNA only granule reaction. Scalebar is 10 μ m. **E.** Sample image of a PCP-8x granule reaction depicting a larger structure. Scale bar is 10 μ m.

We examined the median condensate fluorescence obtained per slncRNA sequence. To get a standardized measurement, we normalized the measured fluorescence values by the number of estimated labeled uracil bases in each sequence (assuming a 35% labeling efficacy as reported by the manufacturer). The standardized quantity is then dependent on the number of molecules in a condensate, as well as on the average fluorescence of a single uracil Atto-488 label. Working under the assumption that the average fluorescence of a single base should not change drastically, the normalized quantity should serve as a rough measure to the number of RNA molecules inside the condensate. The results (Figure 6A) reveal a dependence on slncRNA class, where class I slncRNA molecules yield condensates with weaker fluorescence when compared with class II. The exception to this is PCP-14x/MCP-15x which appears to be weakest on average. This can be due to the reduced fluorescence of a single uracil label brought about by quenching. We further examined the background fluorescence from each slncRNA granule reaction and found roughly similar background levels for all slncRNAs (350-450 [A.U]) (Figure 6B). Normalization of the background (under the assumption that the main contribution to the background is free floating slncRNAs), reveals a dependence on size with the shorter slncRNAs (PCP-4x and PCP-3x/MCP-3x) showing higher normalized fluorescence compared to the negative control (Figure 6C), and longer slncRNAs showing weaker values. Quantification of the area of the localized condensates shows that granule size is generally dependent on slncRNA length rather than on number of hairpins, with shorter slncRNAs resulting in smaller granules (Figure 6D).

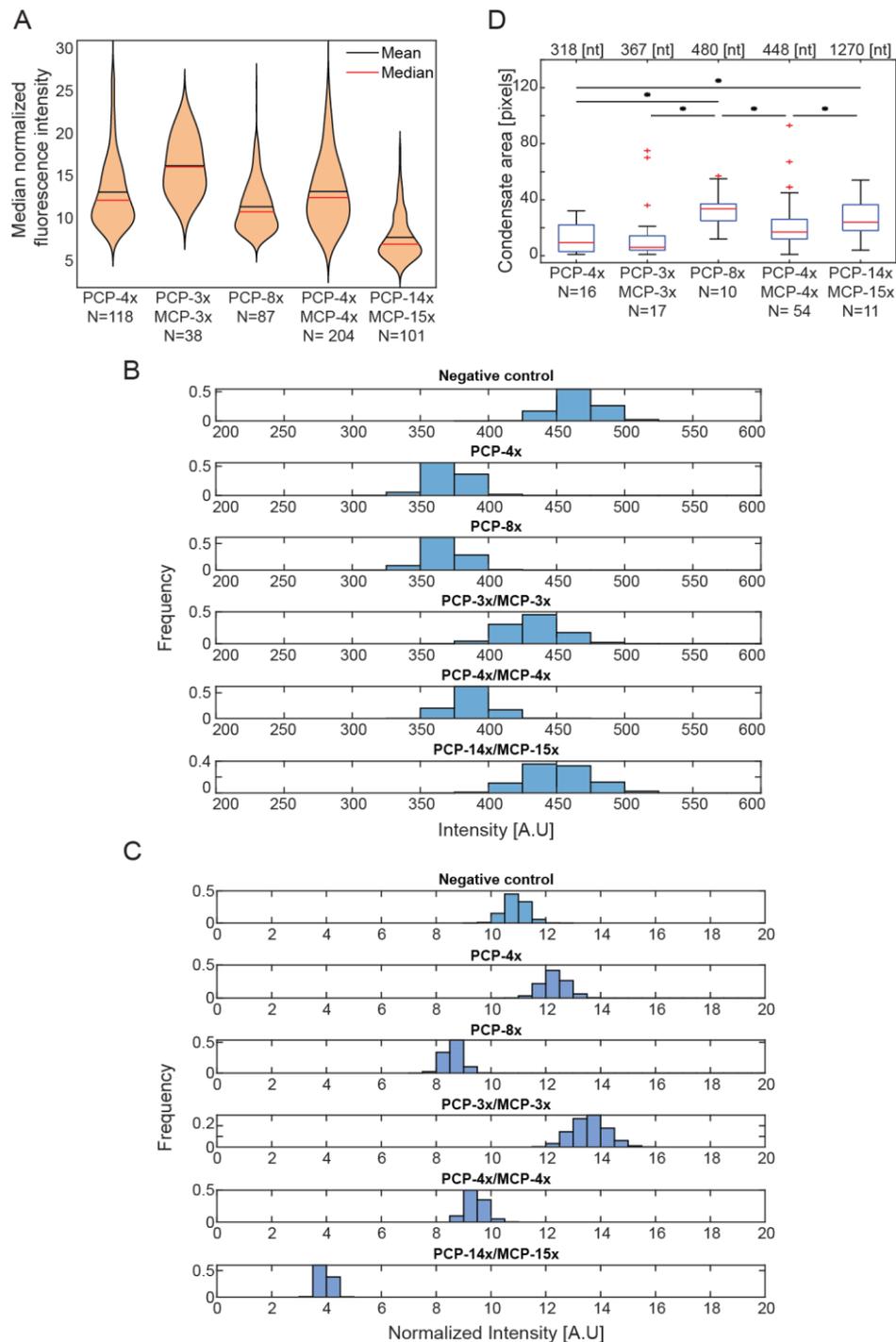


Figure 6 - Further RNA-only granule analysis.

A. Violin plots of median condensate fluorescence of slncRNA-only condensates. **B.** Comparison between the absolute background fluorescence levels from different slncRNA reactions. **C.** Background fluorescence values normalized by the estimated number of labelled uracil bases on each slncRNA molecule. **D.** Boxplots showing area (in pixels) of localized condensates, showing that longer slncRNA molecules typically result in larger condensates. Asterisks denote statistical significance at the 5% level according to a two-sample *t* test. (*p*-values: PCP-4x vs. PCP-8x: $4e-4$; PCP-4x vs. PCP-14x/MCP-14x: 0.046; PCP-4x vs. PCP-3x/MCP-3x: 0.043; PCP-8x vs. PCP-4x/MCP-4x: 0.025; PCP-4x/MCP-4x vs. PCP-14x/MCP-14x: 0.031). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for 'Whisker' corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs.

To further analyze the condensate structure, we fitted the normalized condensate fluorescence intensity distributions to a modified Poisson distribution of the form:

$$p(I) = \frac{\lambda^{\frac{I}{k_0}} e^{-\lambda}}{\left(\frac{I}{k_0}\right)!} \quad (1.1)$$

where I is the observed fluorescence amplitude, λ is the Poisson parameter (rate), and k_0 is a fitting parameter whose value corresponds to the fluorescence intensity associated with a single fluorescent molecule within the burst.

The fits (Figure 7 A-E) reveal three characteristic distributions. For PCP-4x, an exponential distribution is recorded (i.e., $\lambda=0$). For PCP-3x/MCP-3x, and PCP-4x/MCP-4x, a Poisson distribution of $\lambda \sim 1$ seems to be the best fit. Finally, for PCP-8x, and PCP-14x/MCP-15x, a Poisson distribution of $\lambda \sim 2-3$ fit best. These results are consistent with the formation of condensates that are characterized by an increasing number of slncRNA molecules that are cross-linked to form a gel-like "granule", where the number of hairpins encoded into the slncRNA determines the average number of molecules or cross-links within the observed field of granules. Moreover, the interpretation suggested by the shape of the distribution is contrasted by the counter-intuitive observation of decreasing value of the of the fitting parameter K_0 as a function of an increasing number of hairpins (Figure 7F). In this particular context, this observation is manifested by a significantly more gradual increase in mean or median granule fluorescence as compared to what would be naively expected by a simple rescale that takes into account the number of hairpins. Together, these observations suggest that slncRNA granules form via cross-linking interaction of multiple slncRNA molecules, and that an increasing number of hairpins and cross-links lead to a denser condensate. Denser granules, in turn, may result in fluorescence quenching of the labelled uracil bases⁸⁷ leading only to a gradual and disproportionate increase in fluorescence observed.

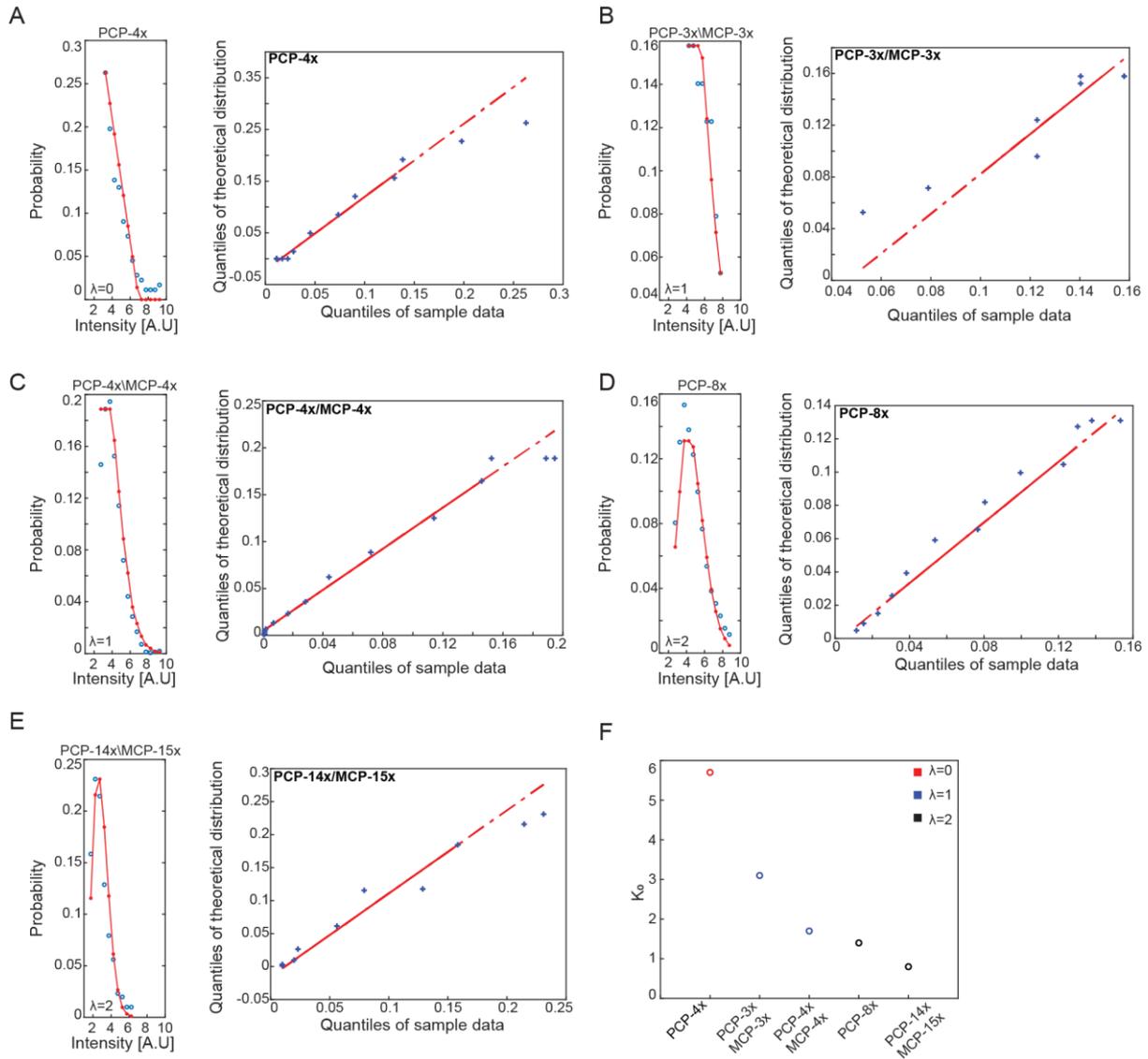


Figure 7 – In vitro RNA-only granules Poisson analysis.

A-E. (Left) Poisson function fits for the median fluorescence intensities of the slncRNA granules. (Right) Quantile-quantile (QQ) plots showing agreement between sample data (experimental observations) and the theoretical Poisson distribution for the fits on the left. **F.** k_0 estimates calculated from the Poisson fits, showing a dependence on the number of binding sites in the slncRNA molecule.

RNA-based granules co-localize with protein-binding partners

To test if the hairpins retain their ability to bind the PP7 phage coat protein while in the granule state, we added recombinant tandem dimer PP7 coat protein fused to mCherry (tdPCP-mCherry) to the granule formation reaction in large excess (reactions were set up with 10-20 nM slncRNA concentration and 800 nM protein concentration) to saturate the slncRNA molecules while accounting for the multiple binding sites present on one slncRNA molecule (Figure 8A). The tdPCP-mCherry version used lacks the necessary moiety to form the wildtype viral capsid (i.e., the FG region)⁷⁴. The images (Figure 8B) show colocalization between the 488 nm channel (Atto-488) and the 585 nm channel (mCherry) for all

slncRNA designs used in the experiment implying that PP7 coat proteins can bind the RNA hairpins in the condensed state. Hence, the slncRNA and their protein partners form synthetic RNA-protein (SRNP) granules. Unexpectedly, PP7-3x granules were witnessed in the presence of the protein, implying that tdPCP-mCherry adds a measure of multivalency to the system, and thus triggers condensation of RNA molecules that do not phase separate on their own at the concentration used. To check that this condensation was hairpin dependent, we tested whether the control RNA (of the same length and GC content as PCP-8x) containing no designed hairpins, condenses either on its own or in the presence of tdPCP-mCherry. In both cases, no condensates were detected in either the 488 nm or 585 nm channels (Figure 8C). Finally, unlike for the slncRNA only case, SRNP granules (particularly for high number of hairpins) show an increased propensity to form large-scale extended structures, suggesting a more complex structure formation and condensation for the SRNP granules as compared with the slncRNA-only case.

To check that the observed colocalization is due to affinity between the coat protein and the corresponding binding sites, we repeated the granule formation reactions with a tandem dimer MS2 coat protein fused to mCherry (tdMCP-mCherry) protein. In this case, granules formed with class I slncRNAs showed no colocalization with the protein, while those formed with class II slncRNAs (which contain MCP binding sites) were colocalized (Figure 8 D, E).

To authenticate the granules as being solid-like RNA-protein structures, we imaged them using structured illumination microscopy (SIM) super resolution microscope with 120 nm resolution. Figure 8F-left shows a sample image of a PCP-14x/MCP-15x granule containing the tdPCP-mCherry protein. The image shows that the slncRNA is found mainly in the periphery of the granule, with filaments protruding into its core, where a high amount of protein is amassed in a network like configuration. The RNA seems to encase the protein cargo in a dense shell-like structure. Figure 8F-right shows a sample image of PCP-4x granules, depicting cage-like structures with a solid protein core and slncRNA filaments protruding and connecting the different structures. RNA-only granules on the other hand appear to be more compact and uniform in nature, akin to solids (Figure 8G).

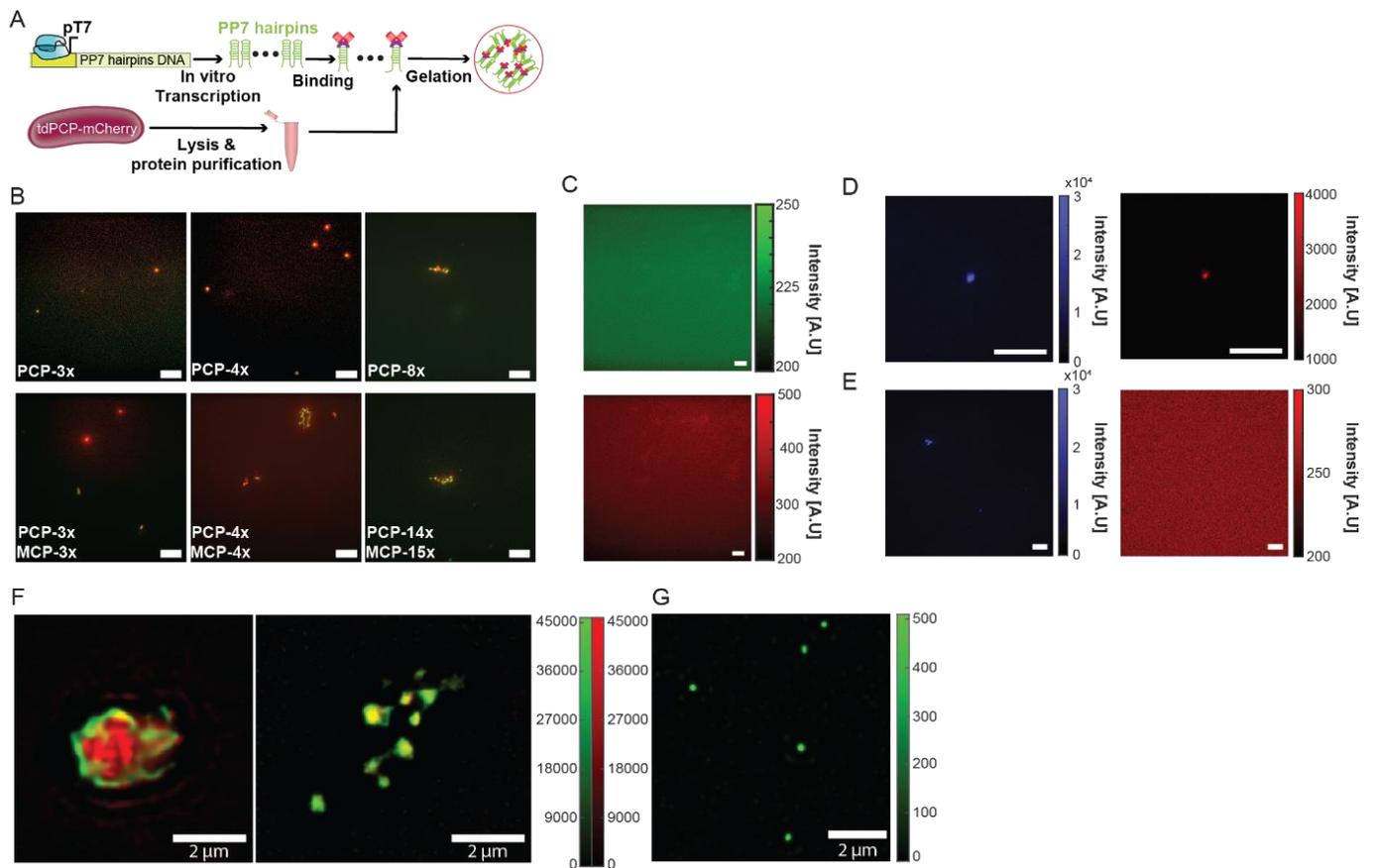


Figure 8 - In vitro RNA-Protein granule microscopy.

A. Construct diagram depicting the suspension of tdPCP-mCherry recombinant protein together with in vitro transcribed slncRNA, resulting in synthetic RNA-protein granules. **B.** Microscopy images showing an overlay of the 585 nm channel (mCherry) and the 488 nm channel (slncRNA). All scale bars are 10 μ m. **C.** Negative control slncRNA-Protein granule reaction. Top – 488 nm channel; Bottom – 585 nm channel. Scalebars are 10 μ m. **D.** Granule reaction with class II slncRNA (PCP-3x/MCP-3x) labelled in AF405 labelling (left), and tdMCP-mCherry (right), showing colocalization between the protein and the RNA. **E.** Granule reaction with class I slncRNA (PCP-8x) labelled in AF405 labelling (left), and tdMCP-mCherry (right), showing RNA structures with no corresponding protein fluorescence. Scalebars are 10 μ m. **F.** Structured illumination super resolution images of (left) PCP-14x/MCP-15x slncRNA-protein granule, and (right) PCP-4x slncRNA-protein granules. Colorbar indicates fluorescence intensity. **G.** Structured illumination super resolution images of PCP-14x/MCP-14x slncRNA-only granules. Scale bar is 2 μ m. Color bar indicates fluorescence intensity.

Next, we measured the median fluorescence intensity of the mCherry protein in different SRNP granules. The distributions of median values (Figure 9A) show a clear dependence on the number of binding sites available for protein binding. First, the PCP-3x/MCP-3x and PCP-4x granules appear to have a similar number of proteins in the granules and are both weaker than PCP-4x/MCP-4x granules, suggesting that PCP-4x slncRNAs inside the granules are not fully occupied by proteins. In addition, the PCP-14x/MCP-15x granules seems to be >2-fold brighter as compared with the PCP-8x granules, despite having <2-fold the number of hairpins. This stands in contrast to the observation that PCP-14x/MCP-15x granules appear to be ~3 times brighter than PCP-4x/MCP-4x granules, reflecting the

difference in the number of binding sites available for binding. Finally, PCP-3x granules appear to be half as bright as PCP-14x/MCP-15x granules, providing more evidence that the former are not RNA-dependent entities. We also observe that when the spacing regions within the slncRNA encode for the MCP hairpins, the formed granules contain a larger protein cargo.

To confirm this observation, we also observed the SRNP granules in the 488 nm channel. Here a slightly more complex image emerges, whereby the median normalized fluorescence values for the RNA granules decline as the number of hairpins increases, hinting once again at a quenching process due to the tightly packed nature of the condensates. In contrast, measurements of RNA-protein granules reveal an increase in fluorescence which is proportional to the number of binding sites available for protein binding. This trend peaks at PCP-8x before the effects of quenching become more dominant for the PCP-14x/MCP-15x. This behavior indicates the existence of an optimum point for slncRNA design in terms of number of binding sites and complexity of the design. (Figure 9B). Together, the observations in both channels indicate that SRNP granules are less dense gel-like structures as compared with the slncRNA-only granules.

Finally, we explored the phase space of SRNP granule formation. To do so, we characterized formation of the PCP-14x/MCP-15x SRNP granules as a function of both slncRNA and protein concentration. For this we produced non-fluorescent RNA molecules (for higher concentrations) and mixed different titers of slncRNA and tdPCP-mCherry protein, each varied over two orders of magnitude. Puncta like structures were detected only for slncRNA and proteins concentrations of 10 nM and 100 nM respectively, or above (Figure 9C). The images display bright puncta that are embedded within a filamentous structure. Quantification of the maximal intensity of the puncta both at time $T=0$ (i.e., beginning of the reaction) and time $T=1$ [hr] (Figure 9D) reveals a fluorescent intensity distribution which declines by two orders of magnitude (i.e., from $\sim 10^5$ to $\sim 10^3$) in a step-like function as the RNA concentration is reduced from 100 to 1 nM, providing further indication that RNA is essential for granule formation. Likewise, the intensity distribution of the puncta declines in a more gradual fashion as the protein concentration is reduced, but overall, a similar disappearance of puncta is observed.

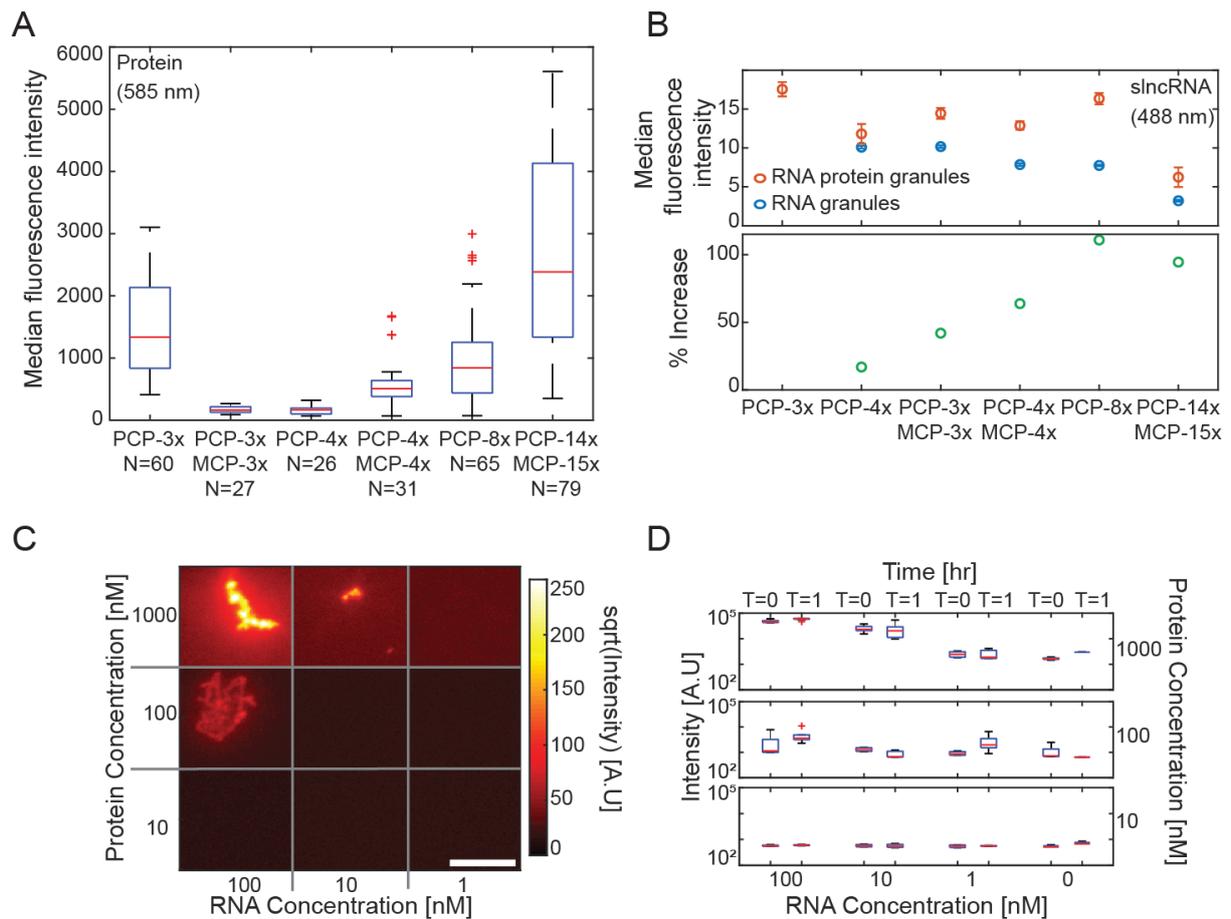


Figure 9 - RNA-Protein granules intensity analysis.

A. Boxplots of median 585 nm (mCherry) fluorescence intensity values collected from multiple granules. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for 'Whisker' corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs. **B.** Top - median 488 nm (Atto488) fluorescence intensity values collected from multiple slncRNA granules (blue) and slncRNA-protein granules (orange). Quenching for the slncRNA granules is empirically estimated at 1x for the PCP-4x and PCP-3x/MCP-3x, 1.34x for the PCP-4x/MCP-4x, 1.37 for the PCP-8x, and 4.2x for PCP-14x/MCP-15x. Note that we assume no quenching for the SRNP granules, except for the case of PCP-14x/MCP-15x. Data presented as median values \pm SE. Bottom—Increase in Atto-488 fluorescence between slncRNA-protein granules and slncRNA only granules, for the different slncRNA molecules. Data in top and bottom panels was collected from: 112 PCP-3x/MCP-3x, 165 PCP-4x, 204 PCP-4x/MCP-4x, 121 PCP-8x, and 89 PCP-14x/MCP-15x, RNA-only granule, and from 91 PCP-3x, 69 PCP-3x/MCP-3x, 30 PCP-4x, 92 PCP-4x/MCP-4x, 85 PCP-8x, and 37 PCP-14x/MCP-15x RNA-protein granules. **C.** Microscopy images for serial dilutions of reaction components taken at T = 1 hr after reaction setup. Highest concentrations show the formation of highly fluorescent filamentous structures, as seen in the top left image. Lower RNA concentrations result in smaller structures, while lower protein concentration result in weaker fluorescence. Scale bar is 10 μ m. Due to high dynamic range, the intensities presented are the square root of the raw data images. **D.** Maximal observed intensity values for each reaction condition at time T=0 and T=1 hr. All distributions were derived from 5 separate microscopy images of granule reaction prepared with the listed concentrations. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for 'Whisker' corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs.

Chapter 2 – Development of an analysis algorithm for gel-like granules

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 that the total fluorescence of the granule consists of three signal components: granule fluorescence, background fluorescence and noise. Granule fluorescence represents the sum of fluorescence intensities of the molecules within the granule core, with occasional addition or subtraction of intensity resulting from molecules entering or exiting the granule. Background fluorescence denotes the fluorescence intensity of the surrounding dilute phase, which contains similar components to the granule albeit at a much lower concentration. Noise here refers to various sources which commonly affect fluorescence microscopy, including shot noise, electronic noise, and dark current noise⁸⁸. The sum of these processes is assumed to be a symmetric, memory-less process (i.e., white noise). We further assume that background fluorescence changes slowly over time, in contrast to granule fluorescence which depends on the dynamic insertion and shedding events occurring in the granule.

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:

$$y(t) = (S(t) + c(t)) \cdot f(t) \quad (2.1)$$

$$c(t) = c_0(t) \cdot f(t) \quad (2.2)$$

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:

$$c_0(t) \approx c_0 = \text{const} \quad (2.3)$$

This leads to:

$$\frac{y(t)}{c(t)} = \frac{S(t) + c_0}{c_0} \quad (2.4)$$

$$S(t) = c_0 \left(\frac{y(t)}{c(t)} \right) - c_0 = c_0 \left(\frac{y(t)}{c(t)} - 1 \right) \quad (2.5)$$

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^{1,89}. In terms of fluorescence intensity, such movement should manifest as a change in signal intensity, lasting over multiple time points and resulting in a higher or lower stable signal level. We term such shifts in the signal as “signal bursts”. To identify such shifts in the baseline fluorescence intensity, we apply a moving-average filter to smooth the data as described in the signal extraction section. This operation biases the fluctuations of the smoothed noisy signal near the bursts towards a gradual increase or decrease in the signal (Figure 10a). Random single fluctuations that do not establish a new baseline level are unlikely to produce a gradual and continuous increase or decrease over multiple time points. We note that this operation alone can mask biologically relevant events that occur on very short timescales, rendering this algorithm less effective for liquid-like condensates 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 10b). For a given trace, the likelihood of observing an instantaneous signal increase event at a specific time-point can be computed as follows

$$P_{inc} = \frac{N(\Delta S(t_i) > 0)}{N_{tot}} \quad (2.6)$$

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:

$$P_{dec} = \frac{N(\Delta S(t_i) < 0)}{N_{tot}} \quad (2.7)$$

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

$$P_{inc}^m = \frac{1}{2^j} \Rightarrow m \log_2(p_{inc}) = -j \Rightarrow m = -\frac{j}{\log_2(p_{inc})} \quad (2.8)$$

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 10c 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), duration (Δt), and time between successive events of the same type (e.g., the time between two positive burst events).

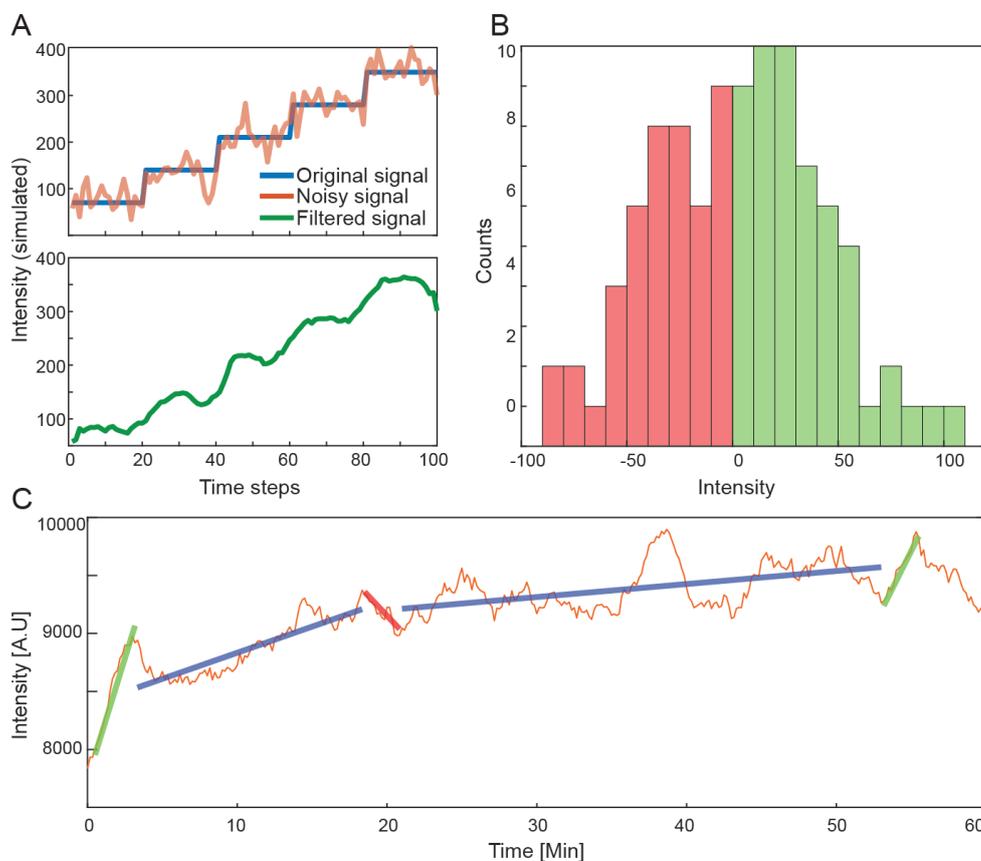


Figure 10 - Identification of burst events.

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:

$$p(I) = \frac{\lambda^{\frac{I}{k_0}} e^{-\lambda}}{\left(\frac{I}{k_0}\right)!} \quad (2.9)$$

where I is the observed fluorescence amplitude, λ is the Poisson parameter (rate), and k_0 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 k_0 values we calculate the theoretical Poisson distribution (Figure 11A) and evaluate its distance from the experimental amplitude distribution in terms of mean squared error (MSE). We then choose the values that minimize the MSE (Figure 11B) as the best fit.

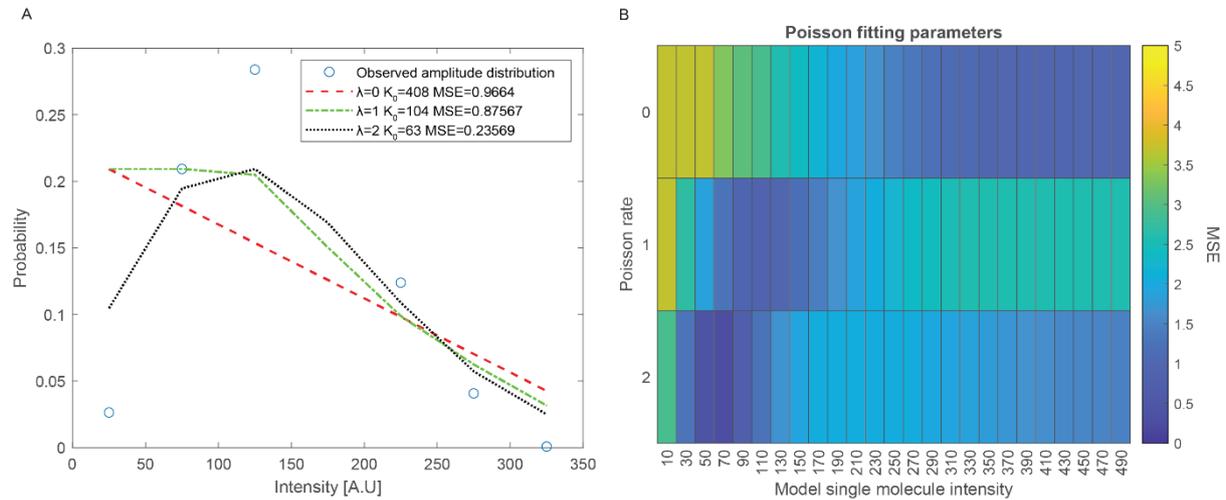


Figure 11 - Poisson fitting example.

A. Sample fitting of an amplitude distribution to three modified Poisson functions, each with a different rate and k_0 parameters. **B.** Heatmap showing MSE values as a function of λ and k_0 values.

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 12).

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 $60 \text{ [A.U]} \times m$ with m being a random number generated from a Poisson distribution with $\lambda=2$ (Figure 12A). The resulting burst amplitudes appear symmetric in nature with relatively similar numbers of positive and negative bursts detected (>1000). Additionally, the range of amplitudes observed is 0-300 [A.U], indicating a significant contribution of the added noise to the underlying amplitudes inserted in the signal (Figure 12B).

Analysis of the constant simulated signals (Figure 12C) reveals once again a symmetric positive and negative amplitude distribution; however, a close examination of the results reveals that the burst amplitude width is smaller by a factor of ~ 5 as compared with the results of the bursty signals (Figure 12D). 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.

For the gradually increasing signals (Figure 12E, 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 behavior profile. A constant signal would correspond to a static observable, akin to a phase separated solid (i.e., glass phase with arrested dynamics). A gradually changing signal would correspond to a constant flux of fluorescent molecules traveling in a single direction. Finally, a bursty signal should best mimic the known behavior of phase separated condensates, corresponding to molecules randomly moving in or out of the dense phase.

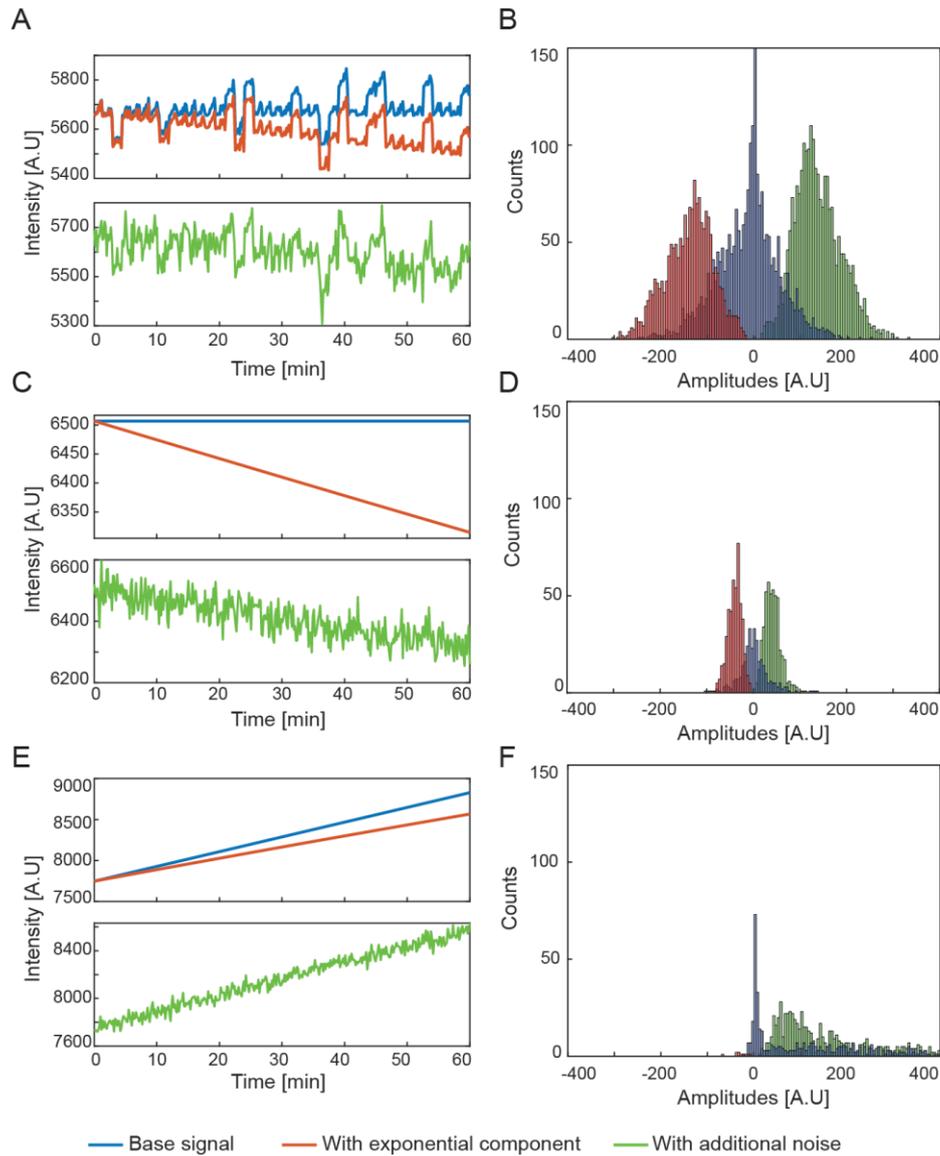


Figure 12- Signal type simulations.

A. Simulated signal with burst events (blue), with photobleaching (orange), and added noise (green). **B.** Amplitude distributions of burst events identified from 1000 bursty signals. **C.** Simulated constant signal (blue), with photobleaching (orange), and added noise (green). **D.** Amplitude distributions of burst events identified from 1000 constant signals. **E.** Simulated signal with slope (blue), with photobleaching (orange), and added noise (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 k_0 , and a random number generated from 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 λ_{sim} values - 1,2,4, and fitted the extracted positive amplitude data to our modified Poisson function.

The positive amplitude distributions for the three cases, demonstrate a visible change in the distributions with higher λ_{sim} values (Figure 13 - Estimating simulation parameters.). The estimation results corresponding to the λ values used in the simulation are also presented (Figure 13 - Estimating simulation parameters. D-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, 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 13 - Estimating simulation parameters.). In the high SNR case, our algorithm correctly estimated $\lambda_{est} = 2, k_{0_{est}} = 223$ (Figure 13I), 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 13J). 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.

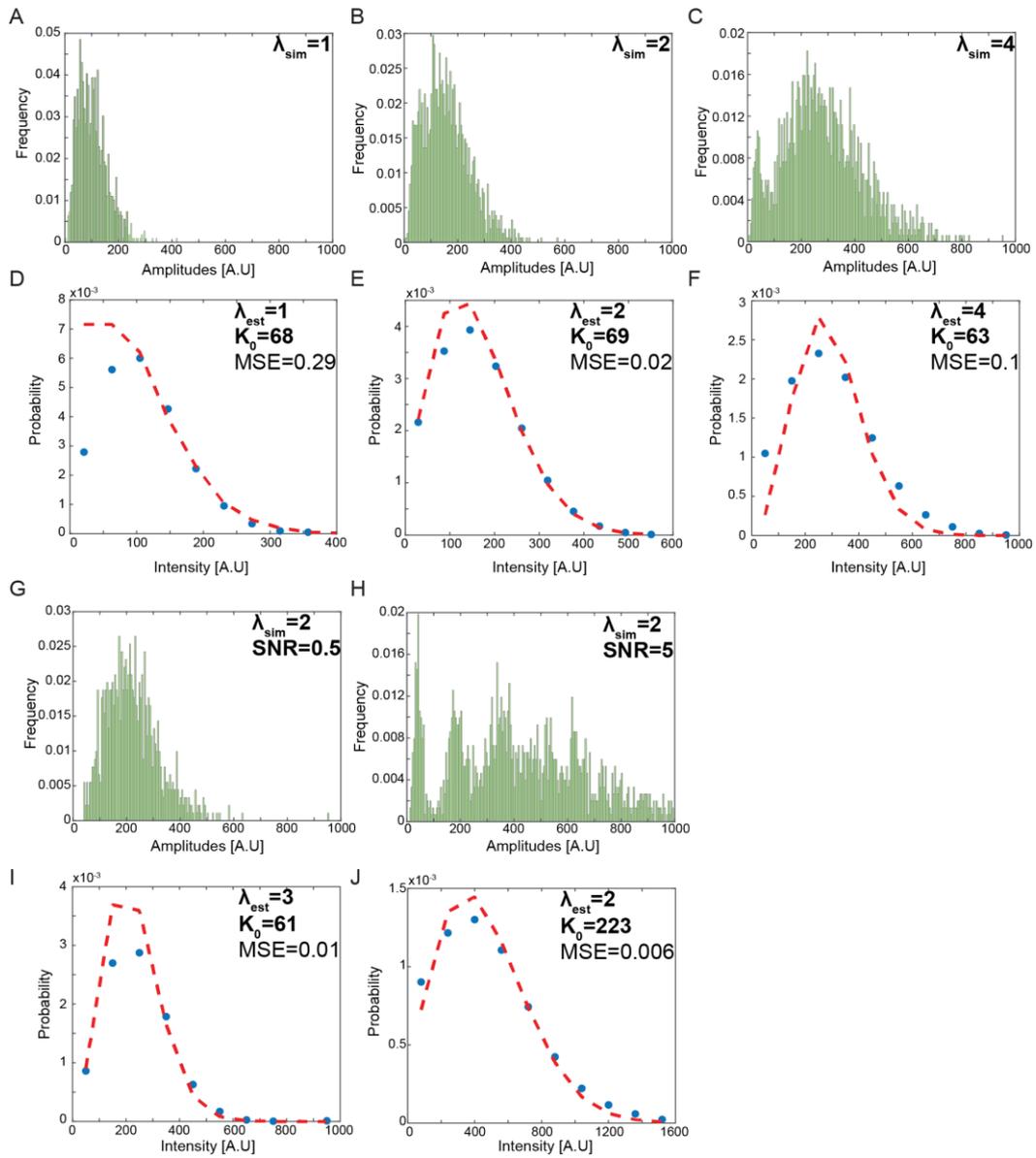


Figure 13 - Estimating simulation parameters.

A-C. Amplitude distributions gathered from 1000 bursty simulated signals with similar base intensity (k_0) and different Poisson parameter (λ_{sim}) values. (A) $\lambda_{sim} = 1$ (B) $\lambda_{sim} = 2$ (C) $\lambda_{sim} = 4$. The distributions demonstrate a visible shift in accordance with the Poisson parameter used in the simulation. **D-F.** Poisson distribution fits of the λ 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 (k_0). **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 (k_0), however in low SNR conditions, the estimation of the Poisson parameter is shifted to $\lambda = 3$.

Chapter 3 – Characterization of synthetic gel-like granules

Dynamic signal analysis of SRNP granules reveals a structure with gel-like characteristics

A hallmark of liquid-liquid phase separation is the exchange of molecules between the dilute phase and the dense phase. This is also true for gels with non-permanent intermolecular interactions, wherein random breaks and rearrangement of the connections which form the inner network allow macromolecules (monomers and small polymers) to diffuse in and out of the gel phase^{39,90–92}, albeit at a significantly slower rate as compared with a high-density liquid phase. These exchange events are predicted to occur independently of one another, at a rate which depends on multiple parameters: the probability of cross linking within the gel network (i.e., number of hairpins), the transient concentration of the molecules in the surrounding solution, and the average diffusion rate of the monomers. The movement of molecules (fluorescent CPs, slncRNA, and CP-bound slncRNA complexes) between the different phases should be reflected by changes in granule fluorescence intensity.

To test whether the synthetic granules display this characteristic we utilized our algorithm described in chapter 2. For this we tracked the fluorescence intensity in both the 488 nm channel (for slncRNA), and the 585 nm channel (for protein), of each granule in a given field-of-view for 60 minutes. As expected, the resulting signals are either decreasing or increasing in overall intensity, and dispersed within them are sharp variations in brightness, that are also either increasing or decreasing (Figure 14A). In Figure 14B we plot the distributions of amplitudes for all three event types, obtained from ~156 signal traces, each gathered from a different granule composed of PCP-14x/MCP-15x and tdPCP-mCherry. We observe a bias towards negative burst or shedding events. Assuming an interpretation that fluorescent burst events correspond to entry and shedding events of slncRNA-CP complexes into or out of the synthetic granules, the amplitude bias towards negative events is consistent with RNA degradation and lack of transcription within the *in vitro* suspension, leading to a net shedding of slncRNA-protein complexes out of the granules over time.

To confirm that we are observing entry and shedding events of what are likely single slncRNA molecules into and out of the fluorescent granules vis-a-vis the signal bursts, we tracked the intensity of PCP-14x/MCP-15x RNA-only granules with and without the presence of RNase A. We first found that at enzyme concentrations above 35 nM, no granules were observed whatsoever, indicating correct activity of the enzyme. At a concentration of 35 nM, we were able to track identified granules for at least 60 minutes. Figure 14C depicts a typical signal of a granule in the presence of RNase, showing a steady decline over several minutes. In addition, while shedding events seem to maintain their amplitude, re-entry events seem to rapidly diminish in amplitude to a median level that is ~10-20% of the original entry burst level (Figure 14D). In particular, the entry burst amplitude reduced in a statistically significant fashion in the latter part of the tracking (30-60') as compared with the first part of the experiment (0-30' – Wilcoxon rank test p -value<0.01). Together, this indicates that slncRNA

degradation occurs outside of the granules, while inside the structure they seem to be protected from degradation, consistent with a gel-like phase.

We then proceeded to gather statistical tracking data for granules produced from all previously described slncRNA designs (including the PP7-3x which does not phase separate on its own at our working concentration). Comparison of the amplitude distributions per design (class I vs. II), (Figure 14E) reveals a dependence on the number of hairpins available for protein binding, where more protein binding sites translate directly into larger amplitudes. As before, PCP-3x is revealed to be an outlier in this case, presenting amplitudes akin to those observed in PCP-4x/MCP-4x granules, providing another indication of a different phase behavior.

In addition, we measured the time duration between events for each granule type. The observed rate (~10 minutes or more) is two orders of magnitude above the typical rate observed in liquid phase separated condensates⁹³ but is in line with the measurements performed on RNA gels³⁹, providing additional confirmation that the SRNP granules are gel-like particles (Figure 14F). Examination of the median time between bursts reveals that shedding events (negative bursts) occur roughly every 10 minutes, regardless of slncRNA design and number of binding sites, indicating a global behavior of the formed granules (Figure 14G-top). Entry events (positive bursts) on the other hand, appear to demonstrate some dependence on the number of binding sites available for protein binding, at least for the slncRNAs with four or more binding sites. For these, the average time between events rises, signifying a reduced ability of bound slncRNA molecules to enter the granules (Figure 14G-bottom). Such a behavior could indicate saturation of the granule, or a high degree of entanglement in the internal granule structure, hindering entrance of new molecules, while also allowing the stochastic shedding of molecules from the periphery of the granule.

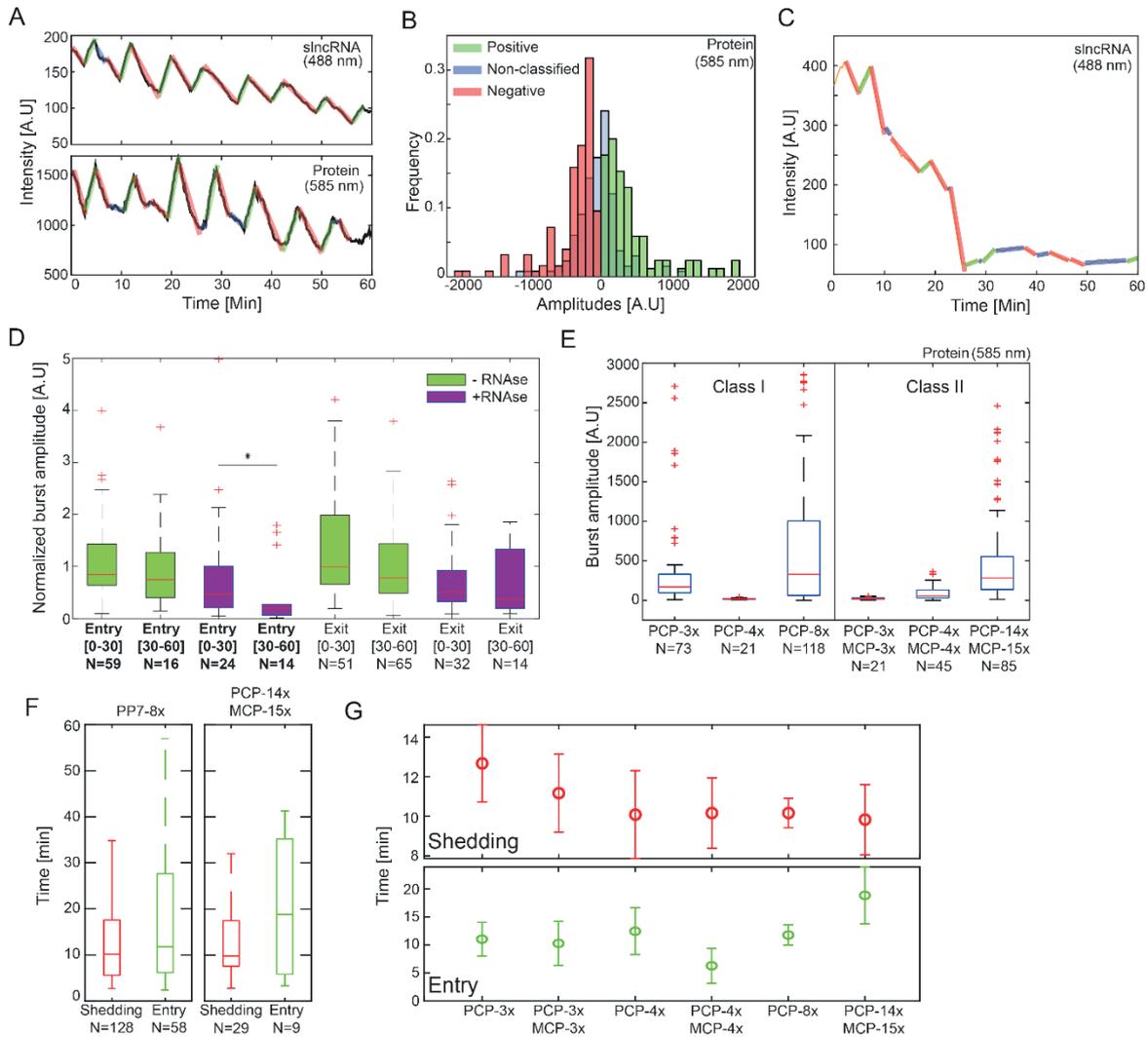


Figure 14 - In vitro dynamic signal analysis.

A. Matching sample traces of both slncRNA fluorescence (top) and protein fluorescence (bottom) measured from a single granule over the course of 60 min. Signals are overlaid with annotations of puncta signal. Annotations represent increasing intensity burst events (green), decreasing intensity burst events (red), and non-classified signal (blue), respectively. **B.** Amplitude distributions gathered from ~156 signal traces in vitro, each collected from a PCP-14x/MCP-15x granule. **C.** Sample signal obtained from the tracking a PCP-14x/MCP-15x granules in the presence of 35 nM RNase A enzyme. **D.** Boxplots showing burst amplitudes measured from PCP-14x/MCP-15x RNA-only granules with and without the presence of RNase A. Amplitudes are normalized by the estimated number of labelled uracil bases in the slncRNA molecule. Asterisk denotes statistical significance at the 5% level according to a Wilcoxon rank sum test (p -value 0.0139). **E.** Boxplots depicting positive amplitude distributions for all slncRNAs. **F.** Sample boxplots depicting distributions of durations between a positive burst and a subsequent positive burst (green), and durations between a negative burst and a subsequent negative burst (red). In (d–f) On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for ‘Whisker’ corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs. **G.** Top—median time between successive shedding events, ($N = 39$ PCP-3x, 30 PCP-3x/MCP-3x, 26 PCP-4x, 45 PCP-4x/MCP-4x, 128 PCP-8x, and 29 PCP-14x/MCP-15x). Bottom—median time between successive entry events, ($N = 23$ PCP-3x, 6 PCP-3x/MCP-3x, 6 PCP-4x, 18 PCP-4x/MCP-4x, 58 PCP-8x, and 9 PCP-14x/MCP-15x). Data in both panels appears as median values \pm SE.

Competition experiments reveal slow granule equilibration times

We next carried out protein exchange or competition experiments on PCP-14x/MCP-15x granules and measured the dynamics and timescales associated with protein replacement. To do so we initially prepared PCP-14x/MCP-15x granules with tdPCP-mCherry and allowed the system to equilibrate. Then, immediately prior to imaging, we added tdPCP-mCerulean and observed the granule dynamics in all three channels (i.e. 405nm for tdPCP-mCerulean, 488 nm for slncRNA, and 585 nm for tdPCP-mCherry). Altogether, we tracked 39 co-labelled granules in multiple experiments. Tracking of the formed granules in both channels (405 nm for cerulean, and 585 nm for mCherry) reveals a variety of signals, indicating various types of mixing of the two labels. For instance, we observed events of apparent displacement where one signal increases as the other decreases (Figure 15A-right) indicating a replacement of the mCherry by mCerulean within the granules. Other signals showed a relatively synchronized signal indicating equilibration of the two protein labels at least within the solution (Figure 15A – left). We next computed the Pearson correlation for each pair of signals and plotted the distribution of correlation coefficients in Figure 15B. The correlation computation shows that while a synchronized signal can be detected in 24 of the 39 signals ($R>0.8$), an anticorrelated or unsynchronized signal is still detected in a significant number of the pairs (15 of 39 signals $R<0.8$) indicating that a significant percentage of the granules are not found in equilibrium.

Comparing the total number of events in both channels during the first 20' reveals that while the number of entry events is similar (Figure 15C- $N=25$ for mCherry and $N=36$ for mCerulean), the number of shedding events is significantly lower in the mCerulean channel (Figure 15D). $N=40$ and $N=2$ for the mCherry and mCerulean channels respectively). At longer tracking durations (20-60'), more mCerulean shedding events are observed ($N=23$) but are nevertheless x3 lower than the total number of mCherry shedding events ($N=69$) over the same duration. Together, the significantly smaller number of shedding events over the one-hour tracking period in the mCerulean channel ($N=109$ for mCherry vs $N=25$ for mCerulean) as compared with the relatively equal number of entry events ($N=69$ for mCherry vs $N=63$ for mCerulean) indicates that while the unbound protein equilibrated as expected, the contents of the granules are still not in equilibrium after 1 hr. This interpretation is supported by time interval measurements between successive burst events (Figure 15E). This analysis reveals a similar rate of entry of tdPCP-cerulean and tdPCP-mCherry proteins into the granules (Figure 15E-left), while a discrepancy between the two rates is observed for the shedding events (Figure 15E-right) in both the duration and total amount of events ($N=73$ for tdPCP-mCherry and $N=4$ for tdPCP-mCerulean, respectively).

We next examined the burst amplitudes as a function of tracking duration intervals (Figure 15 F, G) in the mCherry channel. The data reveals that burst amplitudes for both entry and shedding events decrease over time with a time scale of ~10 minutes. Specifically, burst amplitudes for the mCherry channel are higher in the first 20 minutes as compared with the 20-40- and 40-60-minute time window in the latter

part of the tracking. This indicates a transition from slncRNAs that are fully occupied by tdPCP-mCherry to ones that are increasingly dominated by the tdPCP-mCerulean consistent with rapid equilibration of the proteins within buffer. Consequently, the equilibration of the tdPCP-mCherry and tdPCP-mCerulean within the solution, and the apparent lack of equilibration within the granules at least over the 1 hr duration of the experiments indicated by the unbalanced number of shedding events in both channels, provides additional evidence for a liquid-gel phase transition associated with the SRNP granulation process.

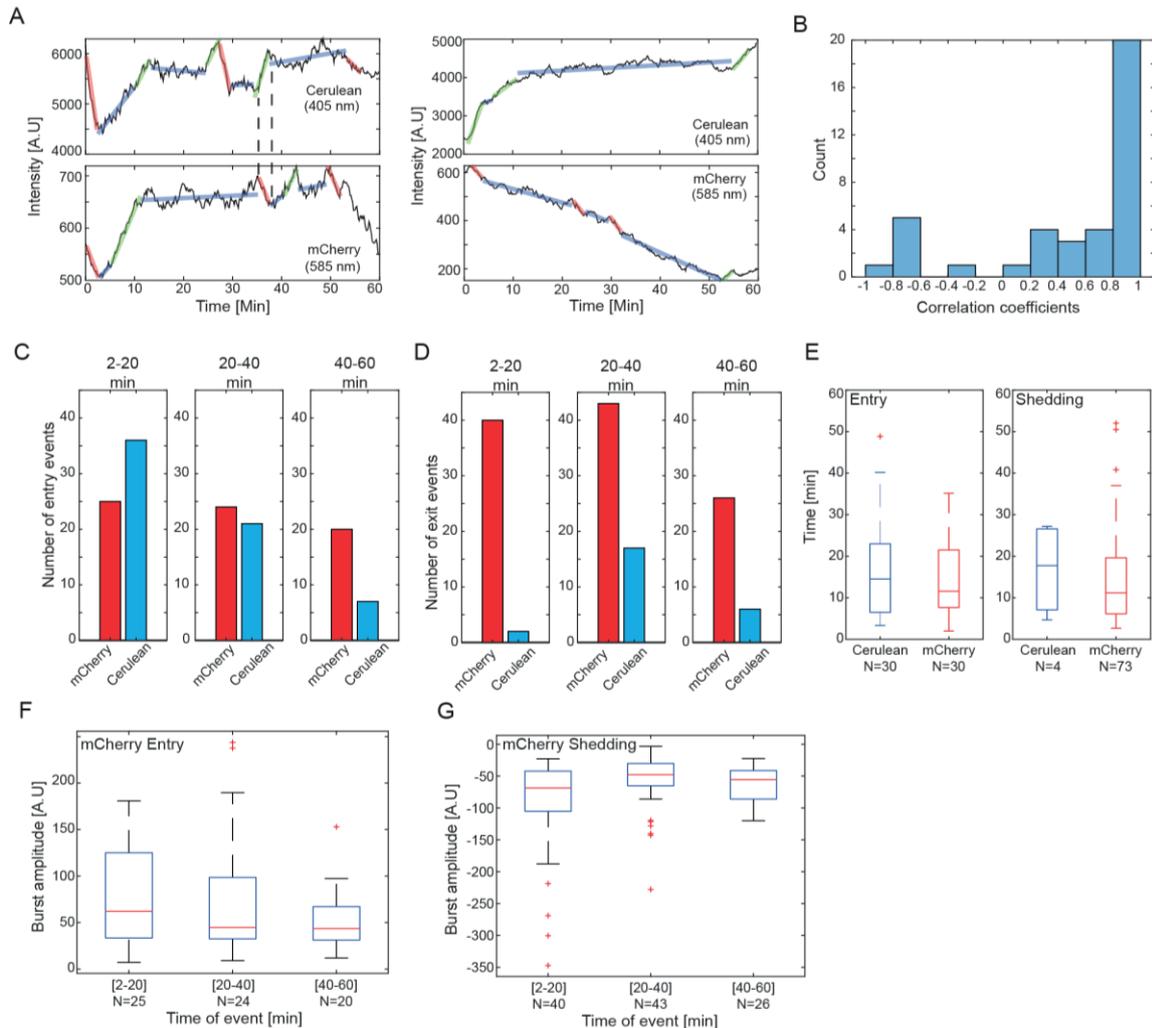


Figure 15 - Competition experiments provide support for liquid-gel phase transition.

A. Sample traces of cerulean fluorescence (top) and mCherry fluorescence (bottom), showing correlation (left) and anti-correlation (right) between the signals. **B.** Pearson correlation distribution plot computed for the 39 co-labelled tracks. Number of entry (**C**) and shedding (**D**) events for both channels for different tracking intervals: (left) 2–20', (center) 20–40', and (right) 40–60'. **E.** Boxplots depicting distributions of durations between an entry event (left), and durations between shedding events (right), for cerulean fluorescence (blue). Burst amplitudes recorded for entry (**F**) and shedding (**G**) events for the mCherry channel as function of the different tracking intervals: (left) 2–20', (center) 20–40', and (right) 40–60'. In (**E–G**) On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for 'Whisker' corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs.

SRNP granules function as protein capacitors

To provide a measure for the number of slncRNA molecules within the granules, we computed the ratio between the mean granule fluorescence and the mean burst amplitude, assuming the average burst corresponds to one slncRNA molecule bound by proteins. The results (Figure 16 A,B) show that type I gel-like granules (PCP-4x and PCP-8x) have a smaller median number of slncRNAs (~5), as compared with type II gel-like granules (~8-10), suggesting that type II granules form better crosslinked structures. For the PCP-3x/MCP-3x, PCP-4x/MCP-4x, and PCP-14x/MCP-15x the ratio in the red channel displays a dependence on the number of hairpins supporting a more robust solid-like behavior when compared with the type I SRNP granules. We next calculated the “net rate of slncRNA loss”, defined as the difference between the total number of observed shedding and entry events, divided by the number of tracked granules (wherein one granule constitutes one hour of tracking data). The rates (Figure 16C) show a difference between class I and class II granules, as well. While the net loss rate for class I granules increases with the number of binding sites, it decreases with the number of binding sites for class II granules. This observation is consistent with type I and type II structures that are characterized by a decreasing and increasing amount of cross-linking, respectively, as a function of the number of hairpins on the slncRNA. Together, this data and the super-resolution microscopy images (Figure 8F) suggest that class I and II granules form different types of gel-like phases, with the former forming a structure that is permeable to proteins while the latter seem to form robust protein storage nanoparticles. In particular, the class II granule characteristics are reminiscent of data and energy storage devices (e.g., capacitors), with the protein cargo replacing the electric charge in the biochemical analog.

To further characterize the “capacitor-like” behavior of the type II granules, we performed a titration experiment with PCP-14x/MCP-15x slncRNAs. We formed granules with a constant slncRNA concentration (120 nM) and different protein concentrations, resulting in a 1:1, 10:1 and 100:1 protein to RNA ratio. We collected shedding burst data for each condition and calculated the previously reported observables. Comparison of the shedding burst amplitudes (Figure 16D) reveals that granules formed with 10:1 and 100:1 ratio have almost identical burst amplitudes, indicating slncRNA binding saturation. A 1:1 protein to RNA ratio results in amplitudes one tenth the intensity, as expected. Interestingly, the increase in burst amplitude also apparently leads to an increase in the number of slncRNA molecules within the granules. This can be seen from the “duration-between-successive-events” distribution (Figure 16E), which shows that the time intervals between entry events in the ratio “1” granules are significantly larger than the ratio “10” and “100” intervals (Wilcoxon p-value <0.005). This combined with a constant rate of shedding independent of ratio leads to granules that are composed of a smaller number of slncRNAs for the ratio “1” as compared with ratio “10” and “100”. Finally, the dependence of the stored protein load on the burst amplitude allows us to define a proportionality constant uniquely for every granule-type. This type of proportionality constant is analogous to

electrostatic “capacitance”, and can be defined by a biochemical analog to the capacitance equation $q=CV$. Here, the biochemical “charge” is the total amount of protein stored within the granules, and the protein concentration within the solution corresponds to a biochemical “voltage”. Consequently, the protein to slncRNA titration measurement provides a more solid footing for the capacitor analogy, suggesting that a potential protein-storage set of applications may be facilitated both in vitro and in vivo.

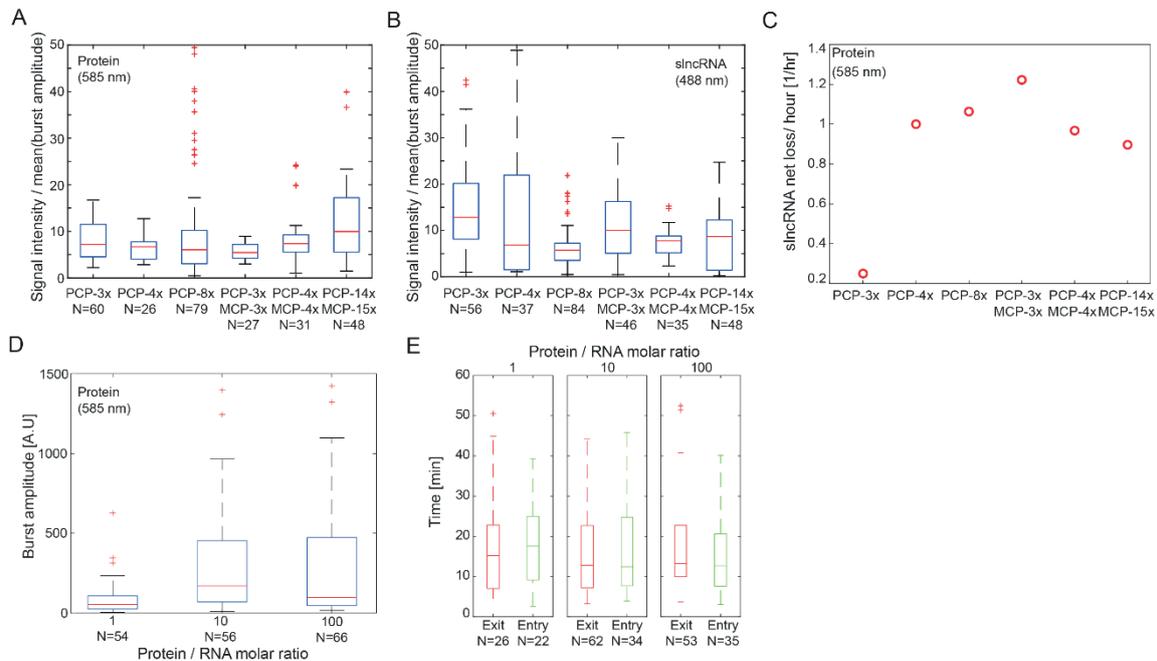


Figure 16 - Granules act as protein capacitors.

A. Boxplots depicting ratio between granule protein fluorescence and median burst amplitude, providing a measure of the protein content inside the granules. **B.** Boxplots depicting ratio between granule slncRNA fluorescence and mean burst amplitude providing a measure of the slncRNA content inside the granules. **C.** Rate of net slncRNA loss for each granule type, showing a difference in trend between class I and class II slncRNA. **D.** Boxplots depicting burst amplitudes measured from granules with different ratios of protein to slncRNA. Granules formed with ratio of 10:1 and 100:1 appear to have similar amplitudes, indicating slncRNA binding saturation. **E.** Boxplots depicting distributions of durations between an entry event (green), and durations between shedding events (red), measured from granules with different ratios of protein to slncRNA. In **(A, B, D, E)** On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for ‘Whisker’ corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs.

Expression of slncRNAs and protein in bacteria yields puncta-like condensates

Given the capacitor analogy, we hypothesized that in vivo granules can be used as devices that store granule-bound proteins indefinitely. This is due to the steady state production of slncRNAs and proteins via cellular transcriptional and translational machinery, that ensures a constant flux of proteins into the granules. To show this, we first proceeded to test whether the granule material characteristics that are

measured in vivo match the in vitro measurements. To do so, we decided to utilize two previously reported slncRNA designs which were shown to yield bright localized puncta in vivo in earlier work⁴⁷. The first slncRNA is of a class II design, PCP-4x/ QCP-5x, consisting of four native PCP binding sites and five native Q β coat protein (QCP) hairpins used as spacers in an interlaced manner. The second slncRNA is the ubiquitous PCP-24x cassette⁹⁴, which from the perspective of this work can be regarded as a class I design slncRNA.

To confirm the formation of condensates in vivo, we encoded the slncRNA component under the control of a T7 promoter, and the tdPCP-mCherry under the control of an inducible pRhIR promoter (Figure 17A). We first wanted to test whether puncta develop in vivo and whether they are dependent on the existence of hairpins in the RNA. For this we co-transformed plasmids encoding either the negative control RNA or the PCP-4x/QCP-5x slncRNA, together with a plasmid encoding for the tdPCP-mCherry protein, into BL21-DE3 E. coli cells. Examination of cells expressing the slncRNA and protein following overnight induction of all components revealed the formation of bright puncta at the cell poles (Figure 17B), which were absent in cells expressing the control RNA which lacks hairpins (Figure 17C). In addition, the difference between the cultures was even visible to the naked eye (Figure 17D), indicating copious amounts of protein which appear to be dependent on number of binding sites encoded in the slncRNA. We believe this phenomenon was missed in the past since such binding sites were exclusively used to track individual mRNA transcripts in vivo where both low concentrations and the effects of translation might hinder the formation of large macro-molecular structures.

Next, to test whether cellular concentration of slncRNA influences the formation of the granules, we quantified the fraction of puncta per cell for cells expressing the PCP-4x/QCP-5x from a multicopy expression vector, and cells expressing the same slncRNA from a bacterial artificial chromosome (BAC) expression vector which is maintained at a single copy level in cells. We found that cells containing the multicopy plasmid frequently present puncta in at least one of the poles, while cells containing the single copy generally show between zero and one punctum (Figure 17E). Given that cells expressing the slncRNAs from single copy vectors still present puncta, we decided to continue using this expression vector in follow-up experiments to reduce variability stemming from copy number differences.

We compared cells expressing the PCP-4x/ QCP-5x or the PCP-24x (expressed from a BAC vector) in terms of the spot per cell fraction. Much like in the in vitro experiments, we found a dependence on the number of binding sites in accordance with the in vitro results and the cross-linking model of gel phase formation^{95,96}. Finally, to test whether the polar localization of the granules is a consequence of nucleoid exclusion⁹⁷, we grew the cells in starvation conditions for several hours, triggering a transition to stationary phase. In stationary phase the nucleoid is known to condense⁹⁸⁻¹⁰⁰, thus increasing the amount of cellular volume which is likely to be molecularly dilute. This, in turn, generates a larger accessible

cellular volume for granule formation, which should lead to different presentation of the phase-separation phenomena as compared with exponentially growing cells. In Figure 17F, we show an image of bacteria displaying ‘bridging’ (the formation of a high intensity streak between the spots) whereby granules seem to fill out the available dilute volume. This behavior is substantially different than the puncta appearing under normal conditions. Such behavior was observed in >40% of the fluorescent cells and was not detected in non-stationary growth conditions. Thus, SRNP granules with characteristics that are consistent with the in vitro observations form in vivo, in a semi-dilute bacterial cytosolic environment and independent of cell-state.

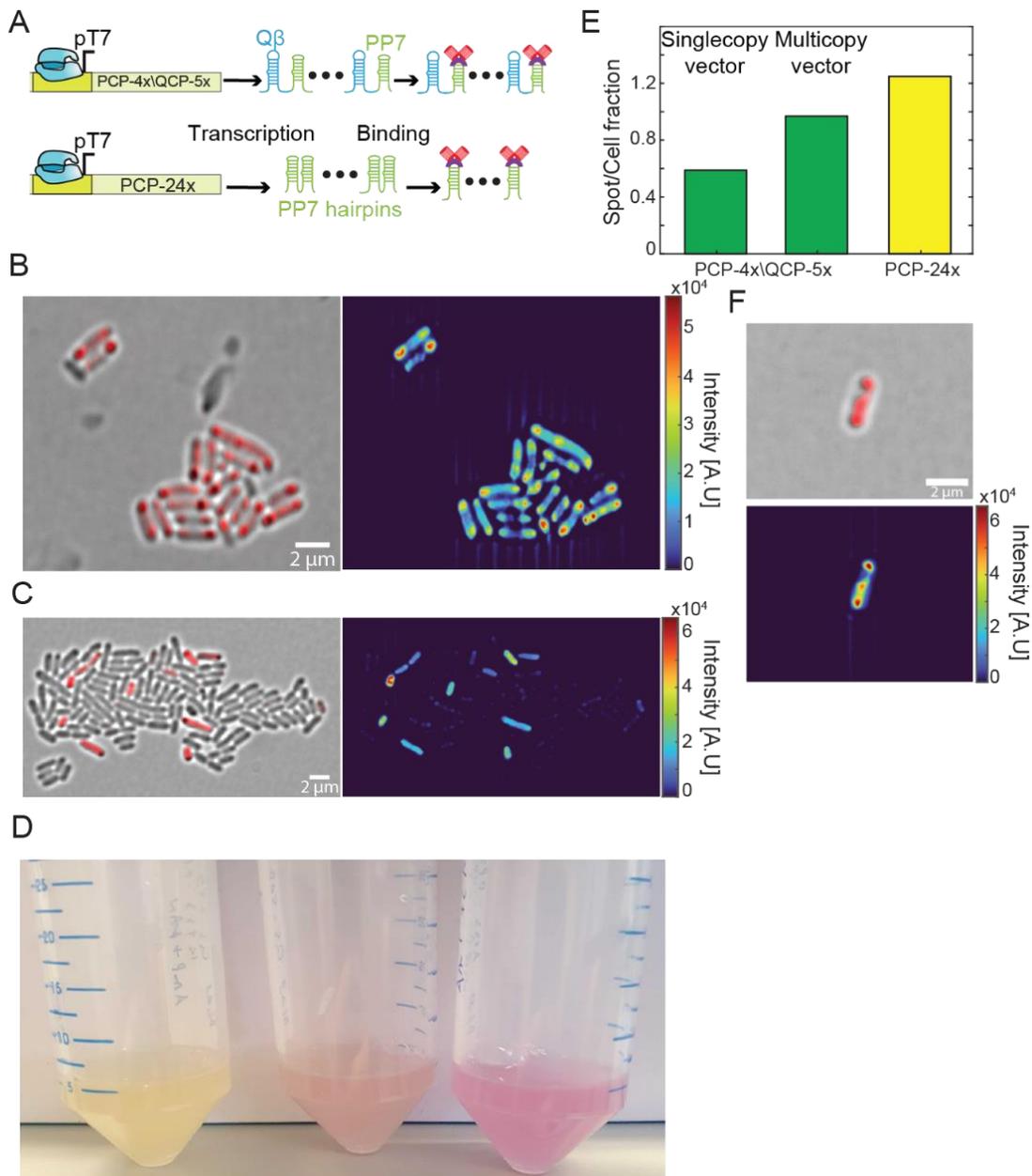


Figure 17 - Synthetic phase separated condensates within bacterial cells.

A. Construct diagram depicting expression of the two slncRNA cassettes used in the in vivo experiments, in the presence of tdPCP-mCherry. **B.** (Left) Merged structured illumination brightfield-585 nm image of cell expressing the PCP-24x slncRNA together with tdPCP-mCherry. (Right) Heatmap of the same image showing a highly fluorescent punctum as cell poles. Color bar indicates fluorescence intensity. **C.** (Left) Merged brightfield-585 nm image of cell expressing the negative control RNA together with tdPCP-mCherry. (Right) Heatmap of the same image showing a weak uniform fluorescence across the cell, color bar indicates fluorescence intensity. **D.** From left to right: E. coli BL21-DE3 expressing tdPP7-mCherry with the negative control slncRNA. E. coli BL21-DE3 expressing tdPP7-mCherry with PCP-4x/QCP-5x slncRNA. E. coli BL21-DE3 expressing tdPP7-mCherry with PCP-24x slncRNA. **E.** Bar plot showing fraction of puncta per cell. (Green) PCP-4x/QCP-5x expressed from either a single copy or a multicopy expression vector. (Yellow) PCP-24x expressed from a single copy vector. **F.** Typical image of fluorescent bacteria in stationary phase, which are different than the 1-2 puncta image obtained for exponentially growing cells. The presented cell shows “bridging” or spreading of puncta. Bottom image show heatmaps of the top image. All scalebars are 2 μm .

slncRNA expression increases cellular protein concentration

To investigate the dynamic properties of granules formed in vivo, we utilized the same analysis approach as was used in the in vitro experiments, with minor differences. Normalizing the fluorescence of the granule by that of the cell for every time point results in a signal vs. time trace largely independent from the effects of photobleaching and cellular background noise, allowing us to search for and measure burst events, as was done previously. In Figure 18 A,B, we plot the amplitude distribution of all three event types (positive, negative, and non-classified), gathered from cells which express the tdPCP-mCherry protein together with a slncRNA, either PCP-4x/QCP-5x (Figure 18A) or PCP-24x (Figure 18B). The symmetry in both shape and spread of the negative and positive distributions indicates that both are measurements of the same type of macromolecule, distinguished only by the direction in which it travels (into or out of the granule). This result contrasts with the in vitro amplitude distribution data (Figure 14B), which presented a skewness towards negative bursts. This implies that in vivo, the transcriptional and translational processes in the cell balance the loss of granule components due to degradation.

Next, we measured the amplitudes of the bursts for both slncRNAs and found that positive and negative amplitudes are proportional to the number of binding sites within the encoded cassette (Figure 18C). In addition, a more quantitative analysis of these distributions (Figure 18 D, E) reveals that a single burst for the 24x cassette is $\sim 2.5\text{-}3\text{x}$ more fluorescent as compared with the 4x cassette, indicating that the molecules transitioning in and out of the 24x granules are slncRNAs partially or fully bound rather than lone proteins. Moreover, estimations of the positive and negative amplitudes are practically equal per slncRNA, providing additional evidence that these are in fact representations of one physical process, with the difference being the directionality of the transitioning slncRNA-protein molecule. Finally, we measured the duration between burst events, revealing that slow shedding and absorption processes on the order of minutes are taking place for the in vivo granules as well (Figure 18F). Altogether, the non-

existence of puncta in cells expressing the negative control RNA, the slow shedding/entry rate of molecules, and the dependence on the number of binding sites, suggest that synthetic RNA protein granules are phase separated condensates *in vivo* and possess the same gel-like characteristics that were observed for the *in vitro* suspensions. Consequently, *in vivo* burst analysis is consistent with the capacitor model, where the amount of protein stored within the SRNP granule seems to be in steady state when there is a steady supply of protein and slncRNA.

Next, to ascertain whether the granules facilitate increased protein titers *in vivo* in accordance with the capacitor model predictions, we measured for each bright granule the mean fluorescence intensity (Figure 18G), and the mean intensity of the cell which contains it (Figure 18H). We observed a dramatic increase in mean cellular fluorescence between cells which express only tdPCP-mCherry and cells which express it together with a slncRNA, suggesting that slncRNA molecules have some effect in the cytosol, regardless of the granules. To quantify this phenomenon more accurately, we measured the total fluorescence of the population using flow cytometry. For this, we grew cells expressing only the protein component (tdPCP-mCherry), and cells expressing both protein and a slncRNA (PCP-4x/QCP-5x or PCP-24x), with different combinations of induction: IPTG (induces the slncRNA) and C4HSL (induces the protein). The data (Figure 18I) shows that cells expressing a slncRNA, regardless of induction (due to T7 leakiness), show higher fluorescence than cells expressing the protein only. In addition, induction of slncRNA expression with IPTG results in an increase in fluorescence, indicating that slncRNA is a deciding factor in this behavior. Finally, cells expressing the PCP-24x slncRNA show higher fluorescence than cells expressing PCP-4x/QCP-5x, displaying a dependence of the cellular protein titer on number of binding sites available for protein binding.

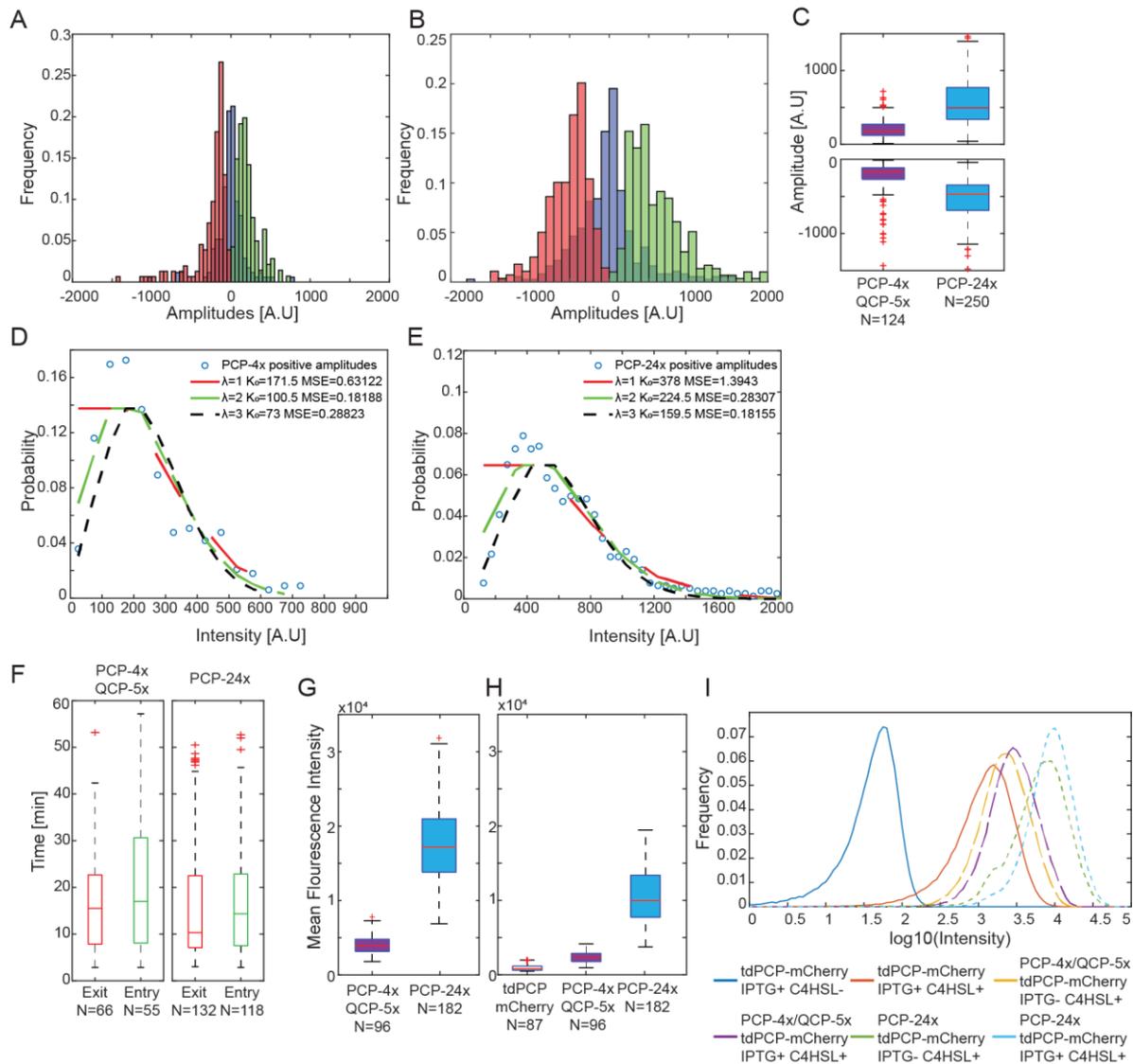


Figure 18 - In vivo granules present similar dynamics as in vitro.

A-B. Empirical amplitude distributions gathered from cells expressing the tdPCP-mCherry protein together with a slncRNA – **(A)** PCP-4x/QCP-5x, and **(B)** PCP-24x. Data obtained from 255 traces **(A)** and 391 traces **(B)**. **C.** Boxplots depicting burst amplitude distributions (top—positive bursts, bottom—negative bursts). **D-E.** Modified Poisson functions fits for the amplitude distribution of insertion events assuming 1, 2, or 3 mean events (λ values). MSE values represent mean squared error between the empirical distribution and the theoretical modified Poisson functions. **(D)** Data collected from 255 PCP-4x/QCP-5x signal traces. **(E)** Data collected from 391 PCP-24x signal traces. **F.** Boxplots depicting distributions of durations between entry events (green), and durations between shedding events (red). **G** Boxplot of mean granule fluorescence intensity. **H** Boxplot of mean cell fluorescence intensity. In **(C, F, G, H)** On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The value for ‘Whisker’ corresponds to ± 1.5 IQR (interquartile rate) and extends to the adjacent value, which is the most extreme data value that is not an outlier. The outliers are plotted individually as plus signs. **I.** Population intensities of *E. coli* BL21 cells expressing tdPCP-mCherry with different slncRNAs and different combinations of induction as measured by flow cytometry.

Chapter 4 – Characterization of gel-like granules with an intrinsically disordered region

RNA with valency of $n=3$ can melt Q β coat protein condensates within bacterial cells

The previously described RNP granule platform utilized the PP7 bacteriophage coat protein (PCP) as the protein component. PCP is characterized by a rigid tertiary structure containing multiple β -strands, rendering it highly ordered and thus lacking the unique characteristics of IDRs⁷⁴. We opted to add an element of disorder while still retaining the simplistic two-component nature of our synthetic granules. Therefore, we decided to utilize the coat protein of the bacteriophage Q β , which contains a documented disordered region¹⁰¹ and is identified as such by state-of-the-art disorder predictors¹⁰² (Figure 19A). We hypothesized that when overexpressed, the Q β coat protein (QCP) would phase separate in cells on its own accord. To test for this, we fused QCP on its C-terminus to an mCherry fluorescent protein, both to tag the coat protein and to render it incapable of spontaneously forming the well-known viral capsid⁷⁵⁻⁷⁸. We cloned the fusion protein into *E. coli* cells under the control of an inducible pRhIr promoter and tested for the formation of bright polar puncta using a fluorescent microscope. Per our assumption, we found bright spots localized to the cell poles for most cells examined (Figure 19B), indicating that the disordered region within QCP provides sufficient multivalency to drive some form of phase separation. To test the effect of hairpin-containing RNA, we designed a synthetic long non-coding RNA (slncRNA) sequence encoding for three hairpins which serve as binding sites for QCP and cloned the corresponding DNA into the *E. coli* BL21 (DE3) strain under the control of a T7 promoter. Observation of cells in which protein expression is induced, but slncRNA expression is not (i.e., basal expression level dependent on the leakiness of the lacUV promoter controlling T7 RNA polymerase expression¹⁰³) reveals homogenous fluorescence across the cells with no puncta whatsoever, indicating that low concentration of hairpin-containing slncRNA can effectively inhibit the phase-separating tendency of QCP (Figure 19C). In contrast, cells in which both slncRNA and protein expression are induced exhibit bright fluorescent puncta localized to the cell poles, indicating that an RNA-dependent condensate has emerged (Figure 19D). Similar expression of a negative control slncRNA (denoted as QCP-Non-Specific (NS)) which does not encode for any hairpins shows that polar puncta are retained, both with and without NS slncRNA induction (Figure 19 E, F).

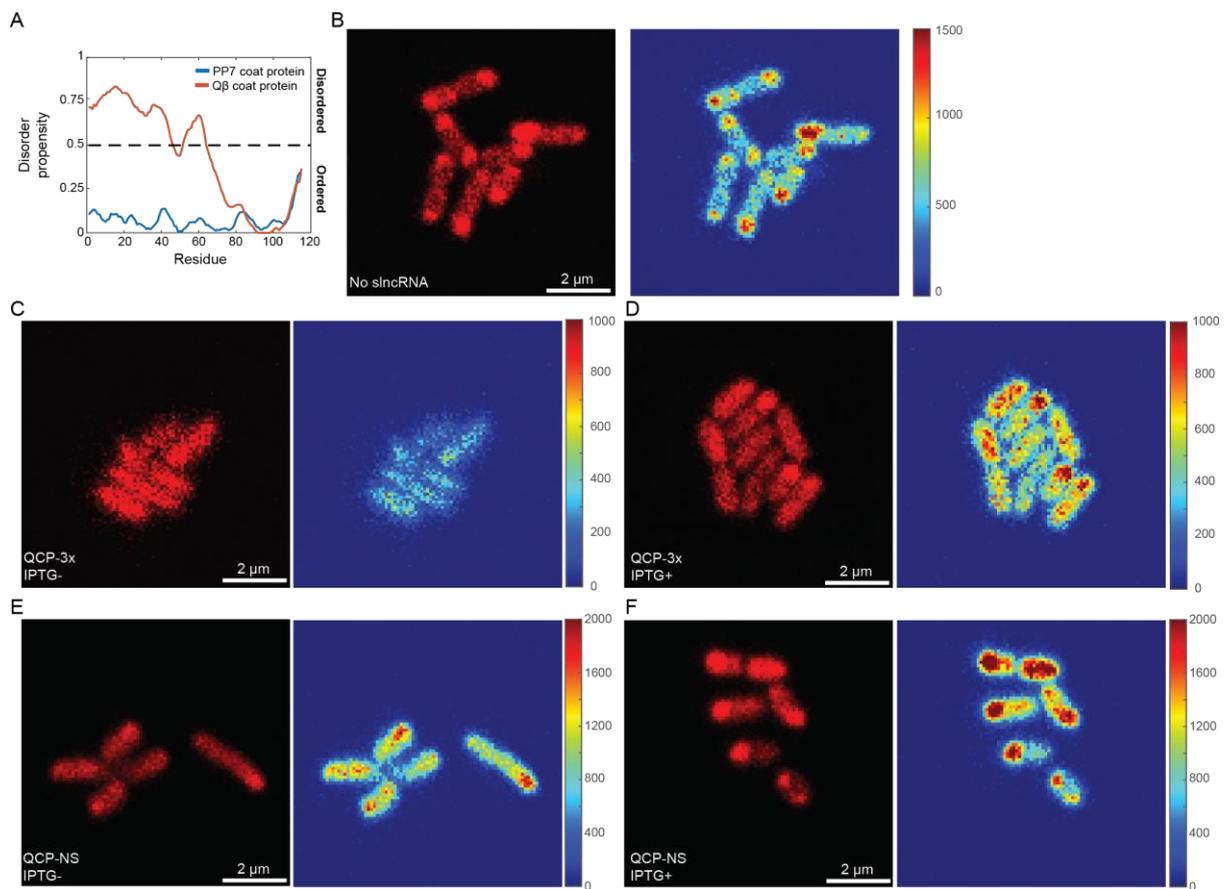


Figure 19 - QCP phase separation behavior in the absence or presence of slncRNAs

A. Disorder propensity predictions of the Q β and PP7 phage coat proteins, in red and blue, respectively. The predictions show a high confidence in the existence of a disordered region in QCP. Prediction by Metapredict. **B.** Sample field of view showing cells expressing just the QCP-mCherry protein with no slncRNA. The cells exhibit bright fluorescent spots localized to the cell poles, reminiscent of phase separated compartments. **C-D.** Sample fields of view showing cells expressing the QCP-mCherry protein together with a slncRNA encoding for three hairpin binding sites, without slncRNA induction (**c**), and with it (**D**). Cells without induction exhibit uniform fluorescence across the cell with no discernible puncta. With slncRNA induction, bright fluorescent spots reappear at the cell poles. **E-F.** Sample fields of view showing cells expressing the QCP-mCherry protein together with NS slncRNA, encoding for no hairpin binding sites, without slncRNA induction (**E**) and with it (**F**). Both cells without NS induction and with it present bright puncta localized to the poles. All scalebars are 5 μ m. Colorbars indicate mCherry fluorescence intensity.

Synthetic RNP granules with Q β coat protein exhibit multiphasic behavior

We hypothesized that RNA valency may be used to modulate the dissolution and regranulation effect. To explore the effect of valency, we designed a set of 5 additional slncRNA sequences, encoding for 1,5,10,12, or 17 hairpins (denoted as QCP- nx with n being the number of hairpins encoded into the RNA sequence), and tested for the appearance of bright spots at the cell poles, with and without slncRNA induction. The results, presented as a fraction of spots per cell (Figure 20A), demonstrate

different behaviors depending on slncRNA induction. Without induction (Figure 20A blue bars), bacteria expressing the QCP-NS slncRNA show the highest fraction of bright spots, as would be expected from a protein-based condensate. slncRNAs with low valency (QCP-1x/3x) appear to be sufficient to melt the protein condensates but cannot phase separate on their own at low (uninduced) concentrations, resulting in virtually no visible spots. Bacteria expressing slncRNAs with higher valency (5 hairpins and above) show an abrupt jump to ratios in the range of 0.4 spots per cell. With induction (Figure 20A red bars), the spot-per-cell ratio increases as a function of valency, starting with an abrupt jump at a valency of three and peaking at a valency of ten, before subsequently decreasing for higher valencies.

Both the abrupt changes from negligible spot-per-cell ratio to cell fraction with a detectible granule population (when increasing from three to five hairpins for uninduced cells, and from one to three in induced cells), and the monotonic increase in cell fraction followed by a monotonic decrease as function of increasing valency number, suggest that a complex phase separation phenomenon is taking place. Specifically, the abrupt change from cell fraction of zero, to a finite cell fraction with granules, supports a co-existence regime between a melted and granular phase that depends on RNA concentration, while the monotonic variation of the cell fraction suggests a more subtle evolution of the granules at higher slncRNA valencies. Consequently, the cell fraction data suggests that more than one phase transition process may be occurring.

To further investigate this behavior, we turned to a fluorescence recovery after photobleaching (FRAP) assay to probe the dynamics of the resulting condensates in the induced slncRNA (i.e. IPTG+) granule state¹⁰⁴. To avoid the effects of condensate aging which could affect the results, we incubated the bacteria for 4 hours with the appropriate inductions (50 μ M C4HSL for protein induction and 0.33 mM IPTG for slncRNA induction) prior to imaging. This duration was determined as optimal for both cell growth and condensate formation. To conduct the FRAP assay, we bleached the bright puncta and measured fluorescence recovery dynamics (Figure 20B). Fluorescence intensity quickly recovered (<15 sec) to some degree for slncRNA variants containing at least three hairpins, indicating a liquid-like behavior, while for the QCP-NS slncRNA lacking hairpins no discernible recovery was observed (Figure 20C). Further examination of the percentage of fluorescence recovery, also known as the mobile fraction, reveals two distinct features (Figure 20D). First, for valency of three and five hairpins, there is a sudden jump in the mobile fraction, from a non-recovered regime to a regime of substantial recovery. Second, at valency of ten hairpins or more, we observe a regime of weak or small mobile fraction. The mobile fraction data suggests a slncRNA-dependent change to the interconnectivity inside the granules

Both the cell fraction and the fluorescence recovery data support a model whereby QCP-granules form via homotypic protein-protein interactions and are either solid-like in nature or are highly interconnected liquid-like condensates, as evident from the lack of fluorescence recovery. Introduction

of low-valency slncRNA to the system (QCP-1x and low concentration of QCP-3x), results in dissolution of the QCP-granules. SlncRNAs with medium valency (high concentration of QCP-3x, and QCP-5x), appear to contribute to the phase separation process, as can be inferred from the increase in spot-per-cell-ratio. Additionally, the increased mobile fraction in this case hints at shorter time scales for internal remodeling (i.e., breaking and reforming of crosslinks between molecules). Together, these two observations imply that slncRNA-protein binding does not compete with protein-protein binding, making the two processes conducive to one another. Finally, higher valency slncRNAs (QCP-10x and above), likely function as scaffolds for the phase transition process with RNA-RNA interactions becoming dominant in the condensate. This implied switch could explain the decrease in mobile fraction, as increasingly interconnected condensates exchange less material with their environment.

To provide further support for this model, we hypothesized that highly interconnected granules (formed with high-valency slncRNAs), may become less restrictive in low slncRNA concentrations, due to RNA-RNA interactions being less dominant. To check this, we carried out the FRAP assay on granules with 10, 12, and 17 hairpins in the absence of inducer. The results (Figure 20E) show a mobile fraction that is either larger, or similar to, what was observed in the fully induced case, providing further support to our interpretation. Together, the data shown in Figure 20 demonstrates our ability to control the internal interconnectivity and remodeling rate of the granules via both the concentration and the number of hairpins (i.e., valency) of the slncRNA.

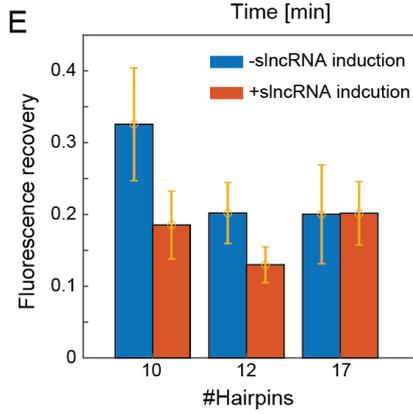
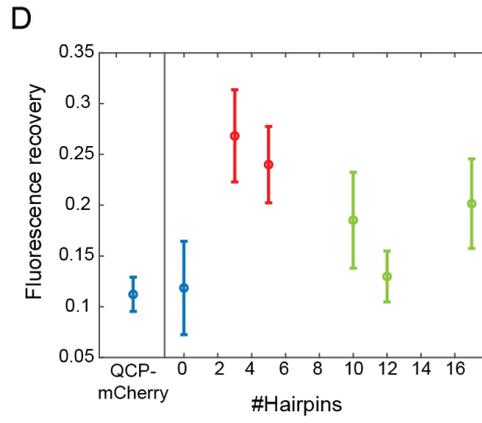
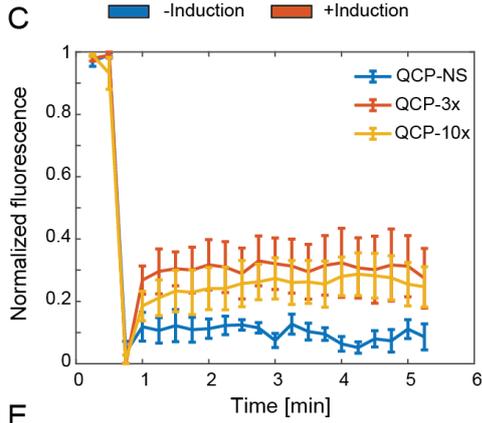
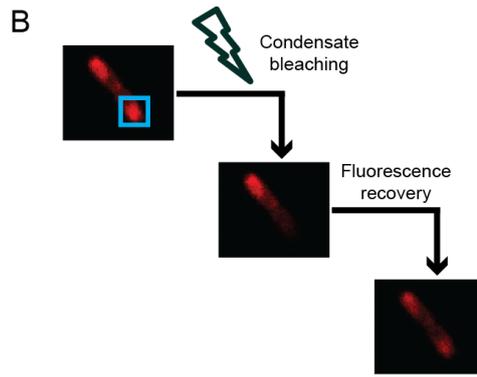
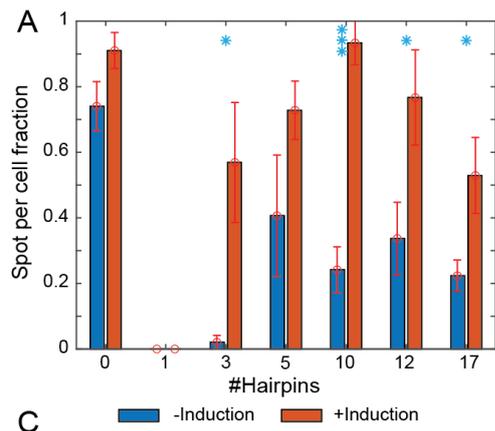


Figure 20 - Fluorescence recovery after photobleaching reveals different regimes.

A. Spot-per-cell ratio for the different slncRNAs, with (red) and without (blue) induction of slncRNA expression. Ratio calculated as the number of fluorescent spots divided by number of cells, aggregated from multiple fields of view. Error bars indicate standard deviation. Stars indicate statistical significance (1 stars for the 0.05 threshold and 3 stars for the 0.001 threshold) according to two-sided t-test (p-values: QCP-3x: 0.03, QCP-10x: 9e-5, QCP-12x: 0.03, QCP-17x: 0.05). **B.** Fluorescence recovery after photobleaching (FRAP) pipeline. Bleaching was localized to a polar condensate, followed by temporal tracking to detect fluorescence recovery. **C.** Sample normalized fluorescence recovery curves measured from cells expressing the QCP-mCherry protein together with QCP-NS (blue), QCP-3x (red), or QCP-10x (yellow) slncRNAs. Curves present averages calculated from 5 different cells. Error bars indicate standard error of the mean. The curves show almost no recovery for QCP-NS, highest recovery for QCP-3x and intermediate for QCP-10x, representing low, intermediate, and high valency slncRNAs. **D.** Fluorescence recovery fractions (mobile fractions) for all tested slncRNAs. Condensates measured from cells which express the negative control slncRNA show similar average recovery to condensates which lack slncRNAs entirely, indicating a protein-dominant phase. Condensates measured from cells which express slncRNAs with 3 or 5 hairpins show highest recovery on average, implying a shift to a less-interconnected condensate with more molecular motion. Condensates measured from cells which express slncRNAs with more than 10 hairpins demonstrate a drop in recovery fraction, indicating a more restrictive environment, probably due to RNA-RNA interactions becoming dominant. **E.** Fluorescence recovery fraction for QCP-10x/12x/17x with and without slncRNA induction. Condensate fluorescence recovers to a higher degree in the absence of induction, indicating more molecular motion. Bars represent mean values. Error bars indicate standard error of the mean. Data calculated from 5 separate cells.

QCP-mCherry titer appears to exhibit a non-monotonic dependence on slncRNA valency

A key feature of our previously reported synthetic RNP granules was their remarkable ability to increase the titer of the associated coat protein fusion inside the cell and form a type of protein-storage vessels, likely as a result of the granules protecting the protein component from cellular degradation processes. This hairpin-dependent increase in titer was visible in liquid culture and was estimated to be about 10-fold via flow cytometry measurements. We hypothesized that a condensate with an RNA-binding protein containing an IDR domain would exhibit a more complex dependence of the protein titer on slncRNA valency. To test this, we measured total cellular fluorescence using flow cytometry of cells expressing QCP-mCherry together with the various slncRNAs, with and without induction of slncRNA expression, as well as cells expressing the mCherry protein alone (i.e., without the phage coat protein) together with the slncRNAs. To allow sufficient time for protein expression and subsequent granulation, we performed this measurement six hours after introduction of the inducing molecules. The results (Figure 21A) demonstrate a substantial increase in fluorescence when comparing the QCP-mCherry fusion to the mCherry protein, regardless of slncRNA induction. This is a result of the fusion to QCP, as was reported in the past⁷⁵. In addition, we see that on average, the fluorescence intensity of cells expressing the mCherry protein is relatively constant regardless of slncRNA valency (Figure 21A–yellow circles), as would be expected from a protein which does not associate with RNA hairpins. For

cells expressing the QCP-mCherry fusion without induction (Figure 21A – red circles), a shift exists between low valency slncRNAs (mean fluorescence of 3.2×10^3 [A.U] for QCP-NC/3x/5x), and high valency slncRNAs (mean fluorescence of 4.1×10^3 [A.U] for QCP-10x/12x/17x). While subtle, this shift is statistically significant, with a p-value of 0.008 using a two-sided t-test. In cells expressing QCP-mCherry with induction (Figure 21 – blue circles), the difference between low and high valency slncRNAs becomes more noticeable (mean fluorescence of 4.8×10^3 [A.U] for low valency slncRNAs, and 7.8×10^3 [A.U] for high valency slncRNAs, with a p-value of 0.025 using two-sided t-test), revealing that high valency slncRNAs lead to increased fluorescence of the associated protein. We note that cells expressing QCP-1x show extremely high fluorescence in both the induced and uninduced cases and were not considered in this analysis.

Interestingly, in cells with slncRNA induction (Figure 21A- blue circles), the measured fluorescence intensities present a non-monotonic relationship to the number of hairpins encoded on the slncRNAs. This is in contrast with the results reported for the PP7 bacteriophage coat protein (chapter 3, Figure 18). To better show this, we plot in Figure 21B the fluorescence intensity normalized by the valency. The strains expressing only mCherry present fluorescence levels that are a monotonic function of valency, as a result of the mathematical operation of normalization. However, cells expressing the various QCP condensates exhibit a shift in the slopes which indicates a multi-exponential dependence on the valency, reflecting a more complex relationship between protein titer and the granule structure.

We aggregated the mean fluorescence intensities based on the three valency regimes of the slncRNA: the sharp exponential decline exhibited for $n=0,1$, a transition region for $n=3,5$, and a more moderate exponential dependence for $n \geq 10$. The data shows a marked difference between mCherry and QCP-mCherry behaviors. QCP-mCherry (Figure 21C - top) fluorescence intensities noticeably decrease when comparing the low and intermediate valency slncRNAs, and increase again for high valency slncRNAs, mirroring the results of the FRAP assay. This is true regardless of slncRNA induction. However, induction leads to higher fluorescence for all tested slncRNAs. In contrast, cells expressing the mCherry protein, which lacks a disordered multivalent protein element, demonstrate a slight increase in intensity as a function slncRNA valency. (Figure 21C - bottom). This observation indicates that slncRNA phase separation by itself may be sufficient to provide some weak measure of titer increase, possibly due to proteins being physically stuck inside the RNA condensate.

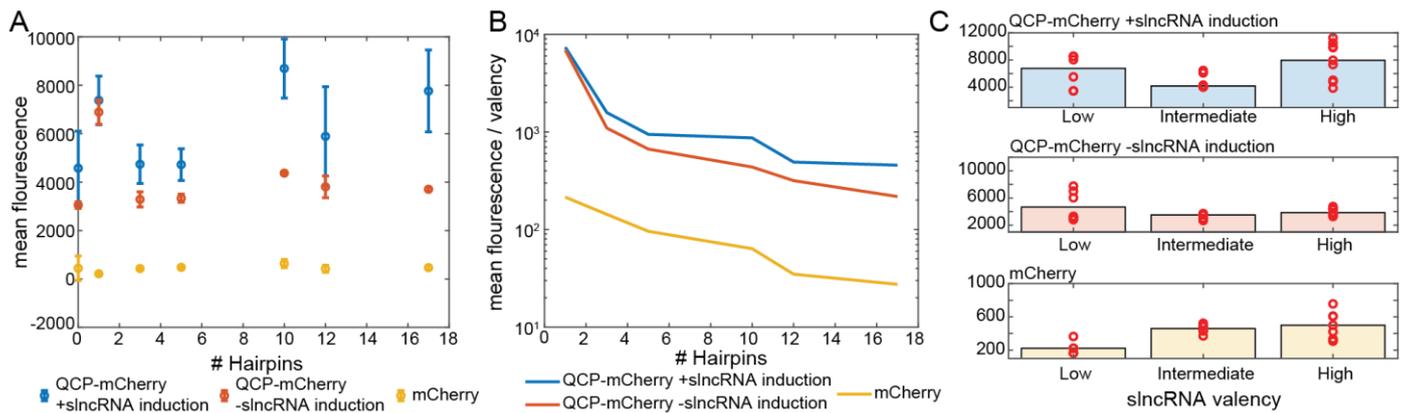


Figure 21 - RNP granule solidity can affect protein titer.

A. Mean mCherry fluorescence flow cytometry measurements from cells expressing slncRNAs together with QCP-mCherry with induction (blue), without it (red), as well as cells expressing slncRNAs together with the mCherry protein (yellow). The measurements show increased expression of QCP-mCherry and a non-monotonic relationship between fluorescence levels and hairpin number on the slncRNAs. Points indicate mean values; error bars indicate standard error of the mean. Values calculated from three biological repeats for all slncRNAs. **B.** Mean fluorescence values from (a), normalized by the valency of the expressed slncRNA. The traces show a monotonic relationship of mCherry in slncRNA valency (yellow trace), whereas QCP-mCherry granules demonstrate breaks, indicative of different behaviors. **C.** Aggregated mean fluorescence measurements per slncRNA valency regime for QCP-mCherry with slncRNA induction (top), QCP-mCherry without slncRNA induction (middle), and mCherry (bottom), showing a dependence of fluorescence level on the granules. Bars indicate average level per phase, circles signify the different data points included (N=6,6,9 for the solid, soft gel, and hard gel phases, respectively).

Multiphasic Q β granules can control protein titers for genes encoded on the slncRNA

Given the behavior of the Q β -based granules, we wondered how the various valency regimes would affect the protein titer of a gene encoded on the slncRNA. To do this, we cloned a ribosome binding site and the gene for a blue fluorescent protein (TagBFP) downstream to our hairpin-encoding DNA sequence (Figure 22A). We first checked for the presence and localization of both the BFP and QCP-mCherry proteins using structured illumination super resolution microscopy and found that QCP-mCherry continued to aggregate at the cell poles while BFP fluorescence was distributed throughout the entire cell (Figure 22B).

We then measured BFP fluorescence via flow cytometry as a function of slncRNA induction, QCP-mCherry induction, and the number of hairpins encoded on the slncRNA. Figure 22 C-F depict sample fluorescence intensity distributions from three slncRNAs: QCP-NS-TagBFP (solid lines), which does not encode any hairpins, QCP-5x-TagBFP (dashed lines), which encodes for 5 RNA hairpins, and QCP-12x-TagBFP (circled lines), which encodes for 12 RNA hairpins, with all four combinations of induction. The results reveal a complicated dependency on slncRNA concentration and valency. With

induction of slncRNA expression, cells containing QCP-12x-TagBFP exhibit the weakest BFP fluorescence, regardless of QCP-mCherry induction (Figure 22 C, D, circled lines). In contrast, cells with QCP-NS-TagBFP and QCP-5x-TagBFP show an increase in BFP fluorescence when QCP-mCherry is induced, demonstrating a dependence on the existence of a phase separated compartment (Figure 22 C, D, solid and dashed lines). In cells without slncRNA induction (Figure 22 E, F), a different pattern emerges, with more hairpins leading to higher BFP fluorescence. This relationship becomes more pronounced when QCP-mCherry expression is induced (Figure 22F), demonstrating the same dependence on the phase separated compartment.

We repeated these fluorescence measurements nine separate times for all slncRNAs and collected mean BFP fluorescence intensities. We then calculated the boost in BFP fluorescence, which we define as the ratio between mean BFP fluorescence intensities of cells with QCP-mCherry induction, and cells without it. It is important to note that virtually all repeats demonstrated a single gaussian distribution, meaning that the entire population of cells (10k per measurement) shifted towards higher or lower fluorescence. In Figure 22G we plot both the fluorescence boost of each individual repeat (black triangles for cells without slncRNA induction and red squares for cells with slncRNA induction) and the average value per slncRNA (black and red bars). The results display a wide range for each slncRNA, with some slncRNAs showing both a fluorescence decrease ($<1x$) and a massive increase in fluorescence ($\sim 100x$) in different biological repeats. This could be a result of internal factors such as sub-optimal slncRNA design, which leads to the protein solid phase outcompeting the slncRNA gel phases (in other terms, the weak protein-protein interactions of the IDR overcome the protein-slncRNA interactions), or external factors such as small differences in oxygen supply due to different positions within the incubator. Of particular note in that respect is QCP-10x-TagBFP, which shows a population of weak boost values and a separate population of extremely high values, both with and without slncRNA induction. This could indicate that a valency of 10 constitutes a bifurcation point in the phase diagram, from the weak-gel phase to the rigid-gel phase. Comparison of the mean value per slncRNA (bars) shows essentially similar values for the +/- induction cases, with the exception of QCP-12x/17x, where cells with high slncRNA concentration show practically no boost in BFP fluorescence, whereas the same cells with low slncRNA concentration present a moderately high increase (Figure 22G). The difference in mean values within these three slncRNAs is statistically significant with student t-test p-values of 0.01, and $7e-4$ for QCP-12x-TagBFP, and QCP-17x-TagBFP, respectively. This constitutes further proof that RNA-RNA interactions can effectively arrest RNA release from the granules, reducing the chance of a ribosome translating the BFP.

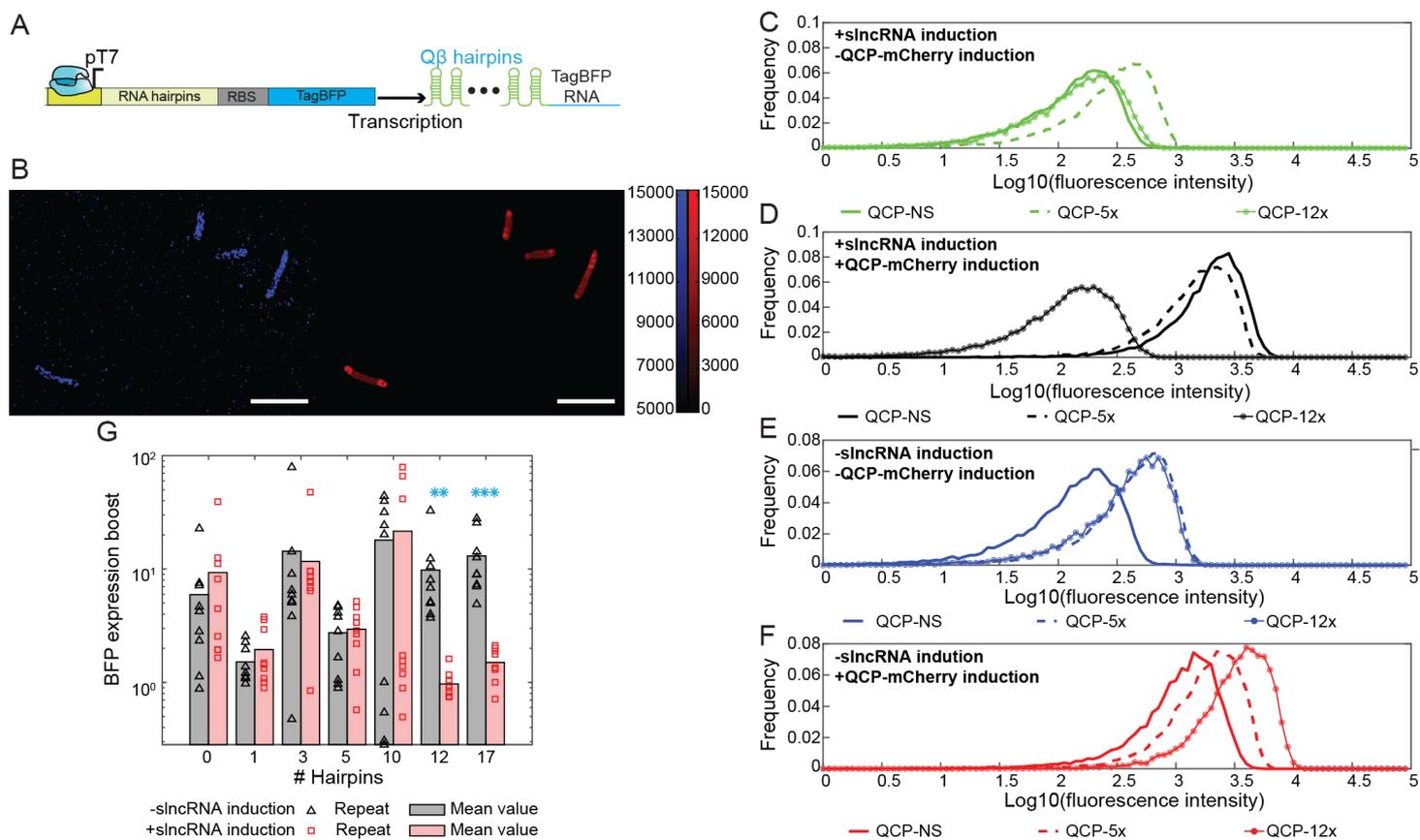


Figure 22 - slncRNA based protein expression leads to significant boost in titer

A. Schematic of the slncRNA-TagBFP construct. **B.** Sample field of view showing cells with both QCP-mCherry (red) and TagBFP (blue) expression. Image captured using a structured illumination super resolution microscope. Scalebars are 5 μ m. Color bars are in arbitrary units. **C-F.** Sample flow cytometry BFP fluorescence measurements of cells expressing QCP-mCherry together with QCP-NS-TagBFP (solid lines), QCP-5x-TagBFP (dashed lines), or QCP-12x-TagBFP (circled lines). **C.** slncRNA induced and QCP-mCherry uninduced. **D.** slncRNA and protein induced. **E.** slncRNA and protein uninduced. **F.** slncRNA uninduced and protein induced. **G.** BFP expression boost defined as the mean BFP fluorescence intensity from cells with QCP-mCherry induction divided by the fluorescence intensity of cells without QCP-mCherry induction, Data is presented as a function of hairpins encoded on the slncRNAs. With slncRNA induction (red) or without it (black). Triangles and squares represent biological repetitions. Bars represent mean values calculated from all nine repeats. Stars represent statistical significance (2 stars for the 0.01 threshold and 3 stars for the 0.001 threshold) of a two-sided t-test (p values: QCP-12x-TagBFP: 0.01, QCP-17x-TagBFP: 7e-4).

Discussion

In this study, we show that synthetic gel-like RNA—protein granules can be designed and assembled using phage coat proteins and RNA molecules that encode multiple CP hairpin binding sites, both in suspension and in vivo. Using fluorescently labeled RNA we show that granule formation is nucleated by RNA-RNA interactions that are proportional to the number of hairpins encoded into the RNA. In addition, the binding of the proteins seems to further enhance and assist the granule formation process.

We then describe and assess a method for investigating the dynamics of solid-like phase separated condensates. Various experimental assays are available to researchers studying phase separation and membraneless organelles. However, these often require specialized equipment or expertise, and many may be inadequate for the study 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 and employs statistical approaches to identify signal segments likely corresponding to molecules entering or exiting the condensate.

Using this novel algorithm, we reveal entry and shedding events of molecules into and out of the granules. By investigating the amplitude of intensity change and rate of occurrence, we show that these events correspond to entry and shedding of protein-bound slncRNA molecules, and that they are dependent on the number of hairpins available for protein binding. Transitioning of macromolecules across a phase boundary is frequently observed in phase-separated condensates, particularly in LLPS-based systems. In particular, the frequency of these transitions reflects the underlying order, internal interactions, and density of the condensed phase. While in liquid-liquid phase separation systems such transitions occur on the scale of seconds or less, here we observe shedding and insertion events on a much longer time scale of minutes or longer, that is more consistent with a solid or gel-like condensed phase. We provide additional evidence for the liquid-gel phase transition underlying the granulation process, by taking the system out of equilibrium and observing the different equilibration times of the dilute solution and dense granule phases. To do so, we used a binding competitor in one experiment and RNase in the other. For those experiments, the liquid phase showed rapid mixing in the former and substantial catalysis in the latter, while the granule phase showed slow mixing and undetectable amount of catalysis, respectively. The out-of-equilibrium experiments allowed us to classify our granules as being gel-like in nature. As a result, we believe that out-of-equilibrium characterizations should be considered a standard tool in similar biomolecular phase separation studies.

We further characterized two options for slncRNA design: a homogeneous design which is comprised of multiple CP hairpin binding site and non-structured spacing regions (class I), and a hybrid design which is comprised of hairpin binding sites and additional hairpins in the spacing regions (class II). We show that the design choice has implications for the granule's protein-carrying capacity and dynamics. In particular, class II granules formed particles with increased cross-linking capability in the RNA-only

granule, which in turn led to an increased ability to insulate the protein cargo in the SRNP granule phase. On the flip side, class I granules were characterized by decreased cross-linking in the RNA-only phase and increased permeability of the protein cargo in the SRNP-granule phase. In addition, class I granules displayed a faster shedding or dissolution rate, which in turn lead to a smaller protein cargo on average. The slow release and strong internal interactions which keep the granules intact for long durations within a gel-like phase, combined with the selectivity of our system due to the RNA binding component, could be utilized as a programmable controlled release mechanism in suitable biological settings. Hence, our granules can be thought of as protein and RNA storage modules akin to a capacitor, with ‘capacitance’ that is dependent on protein concentration, and a monophasic release profile that can be tuned based on slncRNA design. This two-dimensional phase space of capacity vs. rigidity offers substantial flexibility and tunability when designing SRNP granules for a variety of applications.

The capacitor- or storage-like behavior displayed by the SRNP granules implies that *in vivo*, the granules together with the gene-expression machinery form a biochemical analog of an RC-circuit. In a conventional RC-circuit, energy is stored within the capacitor for release at a later time. Such circuits are often used to protect electrical devices against sudden surges or stoppages of power. Here, the protein and slncRNA flux into the cytosol correspond to the current, which results in the formation of fully “charged” SRNP granules. This genetically encoded slncRNA and protein storage facility, which is constantly maintained, effectively increases the protein and slncRNA content of the cell beyond the steady-state levels facilitated by standard transcription, translation, RNA degradation, and proteolysis. This storage capacity is precisely the function that is carried out by capacitors in RC-circuits, allowing electrical devices to function even after “power” is cut-off.

We further studied the phase-separation behavior of more complex synthetic RNA-protein granules, where the protein component (Q β phage coat protein - QCP) contains both a specific RNA binding moiety and a known intrinsically disordered region (IDR). We characterized the QCP RNP granules *in vivo* as a function of slncRNA valency (i.e., number of QCP binding sites) using cell fraction analysis, QCP-fluorescence intensity measurements, FRAP analysis, and via fluorescence measurement for a BFP gene encoded on the slncRNA. Our results reveal a complex behavior dependent on slncRNA valency. Specifically, the observed dynamics of granules in cells expressing just the QCP-mCherry protein, as well as cells expressing the QCP-NS slncRNA which encodes no hairpins, is suggestive of a highly interconnected granule composed mainly of proteins. Expression of a slncRNA with a valency of one, leads to granule dissolution. In cells which express intermediate valency slncRNAs (n=3/5), we notice the reappearance of granules characterized by a high mobile fraction (2.5-fold higher), pointing to an enhancement of the phase transition process. Finally, cells expressing slncRNAs with high valency (n=10/12/17), present granules with a reduced mobile fraction, indicative of a more interconnected granule stemming from the increased RNA-RNA interactions.

Extrapolating from this data and applying the principles of polyphasic linkage theory⁶⁰, we can construct a theoretical behavior profile dependent on RNA valency and concentration. Non-valent RNA such as our QCP-NS results in granules dominated by homotypic protein-protein interactions, regardless of RNA concentration, implying that the Q β coat protein acts as a scaffold molecule which undergoes phase separation. These granules are abundant in the cells, and exchange little to no material with the surrounding cellular environment. The presence of these protein granules boosts both QCP-mCherry and TagBFP titers, suggesting an influence of the granules on the RNA despite no apparent association between them. A possible explanation for this is non-specific interactions between the RNA component and the periphery of the already-formed protein condensates.

At intermediate valencies, we observe a regime with virtually no granules in the cells (i.e., the absence of spots for QCP-1x), which we refer to as the ‘melted’ state. This phenomenon can be explained using the polyphasic linkage model, where the proteins act as scaffold molecules which drive phase separation, and slncRNAs serve as ligands. QCP-1x, which can be considered a monovalent ligand, suppresses scaffold phase separation by either directly competing with the protein-protein interaction capacity of the coat proteins, or by enhancing their excluded volume⁶⁰. In this regime, QCP-mCherry expression remains high but BFP expression does not increase significantly (boost of ~1x).

A further increase in the valency enables the formation of homotypic RNA-RNA interactions, and heterotypic protein-RNA interactions¹⁰⁵. Here, the increase in spot-per-cell fraction indicates that protein-RNA interactions do not interfere with protein-protein interactions, thus contributing to the forces driving phase separation. According to the polyphasic linkage model, these intermediate valency slncRNAs function as multivalent ligands which bind to scaffold regions that do not actively promote phase separation (i.e., spacers). This augmented phase separation process leads to a more ‘relaxed’ internal environment which enables maximal exchange of molecules with the dilute phase (i.e., increased mobile fraction).

Finally, at high valency and high slncRNA concentration, the granules become dominated by RNA-RNA interactions, leading to a decrease in both the ratio of granules in the cells and the exchange of molecules with the environment. The most prominent features of this regime are the significant drop in TagBFP expression and a decreased mobile fraction.

These results indicate that the QCP granule platform may form the basis for a promising avenue for increasing the efficacy of protein production within bacteria. Specifically, our findings suggest two different approaches. The first relies on fusing the target protein to QCP. Here, we found a maximal boost of ~30x on average for mCherry fluorescence using the n=1 valency slncRNA as compared with the case for mCherry alone. The second approach is to encode a gene directly downstream from the protein binding moiety. This approach yielded a boost of up to x80 in BFP expression for the n=10 valency. Nevertheless, several challenges remain when considering biotechnological applications using

either approach. Notably, for the first approach the protein mass is potentially trapped inside a granule in which the correct folding of proteins is possible, but not guaranteed. Furthermore, efficient extraction of the protein from the granules might require heating or use of other denaturing agents, potentially damaging or destroying the target protein in the process. Finally, the extracted protein is a fusion between the desired protein and QCP, requiring protease cleavage and a subsequent purification step to obtain a final product. For the second approach, a lack of reproducibility was observed, where some biological repeats yielded a large boost factor, and others of the same system exhibited a much different result, with particularly sharp dichotomy observed for the $n=10$ valency slncRNA. Consequently, additional study is needed in order to convert these findings to a robust biotechnological application.

To the best of our knowledge, this work constitutes a first-of-its-kind completely synthetic model of naturally occurring RNP granules in which both RNA and protein components not only contribute to the formation of the granules but are also capable of changing internal granule features and dynamics. We have demonstrated that RNA valency with respect to the protein can overcome protein-protein interactions mediated by intrinsically disordered regions and lead to a nearly complete dissolution of the protein granules in low valency RNA, or to the formation of an RNA-dominant granule. The concept of IDR modulation via RNA, as demonstrated here, could be a promising therapeutic strategy for pathological conditions related to changes in IDR behavior¹⁰⁶. Finally, the synthetic system presented here could be a useful tool to further explore multiphasic condensate formation provided that naturally occurring IDRs are included enabling a more controlled exploration of complex phase behaviors.

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* **24**, 288–304 (2023).
2. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J Biol Chem* **294**, 7115–7127 (2019).
3. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology* **30**, 39–58 (2014).
4. Jin, X. *et al.* Membraneless organelles formed by liquid-liquid phase separation increase bacterial fitness. *Science Advances* **7**, eabh2929 (2021).
5. Boeynaems, S. *et al.* Phase Separation in Biology and Disease; Current Perspectives and Open Questions. *J Mol Biol* **435**, 167971 (2023).
6. Ganser, L. R. & Myong, S. Methods to Study Phase-Separated Condensates and the Underlying Molecular Interactions. *Trends in Biochemical Sciences* **45**, 1004–1005 (2020).
7. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
8. Dai, Y., You, L. & Chilkoti, A. Engineering synthetic biomolecular condensates. *Nat Rev Bioeng* **1**, 466–480 (2023).
9. Alberti, S. *et al.* A User's Guide for Phase Separation Assays with Purified Proteins. *J Mol Biol* **430**, 4806–4820 (2018).
10. Zhang, Y. & Shen, L. An in vitro Assay to Probe the Formation of Biomolecular Condensates. *Bio Protoc* **13**, e4813 (2023).
11. Bracha, D., Walls, M. T. & Brangwynne, C. P. Probing and engineering liquid-phase organelles. *Nat Biotechnol* **37**, 1435–1445 (2019).
12. Rhine, K., Skanchy, S. & Myong, S. Single-molecule and ensemble methods to probe RNP nucleation and condensate properties. *Methods* **197**, 74–81 (2022).

13. Orti, F., Navarro, A. M., Rabinovich, A., Wodak, S. J. & Marino-Buslje, C. Insight into membraneless organelles and their associated proteins: Drivers, Clients and Regulators. *Computational and Structural Biotechnology Journal* **19**, 3964–3977 (2021).
14. Ryan, V. H. & Fawzi, N. L. Physiological, pathological, and targetable membraneless organelles in neurons. *Trends Neurosci* **42**, 693–708 (2019).
15. An, H., de Meritens, C. R. & Shelkovernikova, T. A. Connecting the “dots”: RNP granule network in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1868**, 119058 (2021).
16. Wang, B. *et al.* Liquid–liquid phase separation in human health and diseases. *Sig Transduct Target Ther* **6**, 1–16 (2021).
17. Ripin, N. & Parker, R. Formation, function, and pathology of RNP granules. *Cell* **186**, 4737–4756 (2023).
18. Buchan, J. R. mRNP granules: Assembly, function, and connections with disease. *RNA Biology* **11**, 1019–1030 (2014).
19. Tauber, D., Tauber, G. & Parker, R. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends in Biochemical Sciences* **45**, 764–778 (2020).
20. Pushpalatha, K. V., Solyga, M., Nakamura, A. & Besse, F. RNP components condense into repressive RNP granules in the aging brain. *Nat Commun* **13**, 2782 (2022).
21. Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat Rev Mol Cell Biol* **10**, 430–436 (2009).
22. Al-Husini, N. *et al.* BR-Bodies Provide Selectively Permeable Condensates that Stimulate mRNA Decay and Prevent Release of Decay Intermediates. *Molecular Cell* **78**, 670-682.e8 (2020).
23. Banerjee, P. R., Milin, A. N., Moosa, M. M., Onuchic, P. L. & Deniz, A. A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angewandte Chemie International Edition* **56**, 11354–11359 (2017).

24. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
25. Langdon, E. M. *et al.* mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).
26. Zhang, H. *et al.* RNA Controls PolyQ Protein Phase Transitions. *Molecular Cell* **60**, 220–230 (2015).
27. Schwartz, J. C., Wang, X., Podell, E. R. & Cech, T. R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Reports* **5**, 918–925 (2013).
28. Fernandes, N. & Buchan, J. R. RPS28B mRNA acts as a scaffold promoting cis-translational interaction of proteins driving P-body assembly. *Nucleic Acids Research* **48**, 6265–6279 (2020).
29. Protter, D. S. W. *et al.* Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Reports* **22**, 1401–1412 (2018).
30. Mittag, T. & Parker, R. Multiple Modes of Protein–Protein Interactions Promote RNP Granule Assembly. *Journal of Molecular Biology* **430**, 4636–4649 (2018).
31. Jonas, S. & Izaurralde, E. The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev.* **27**, 2628–2641 (2013).
32. Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology* **44**, 18–30 (2017).
33. Chakrabarti, P. & Chakravarty, D. Intrinsically disordered proteins/regions and insight into their biomolecular interactions. *Biophysical Chemistry* **283**, 106769 (2022).
34. Uversky, V. N. Intrinsic Disorder, Protein-Protein Interactions, and Disease. *Adv Protein Chem Struct Biol* **110**, 85–121 (2018).
35. Lasker, K. *et al.* The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nat Commun* **13**, 5643 (2022).

36. Garaizar, A., Sanchez-Burgos, I., Collepardo-Guevara, R. & Espinosa, J. R. Expansion of Intrinsically Disordered Proteins Increases the Range of Stability of Liquid–Liquid Phase Separation. *Molecules* **25**, 4705 (2020).
37. Garcia Quiroz, F. *et al.* Intrinsically disordered proteins access a range of hysteretic phase separation behaviors. *Science Advances* **5**, eaax5177 (2019).
38. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
39. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
40. Krzyzosiak, W. J. *et al.* Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Research* **40**, 11–26 (2012).
41. Bertrand, E. *et al.* Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell* **2**, 437–445 (1998).
42. Tutucci, E. *et al.* An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat Methods* **15**, 81–89 (2018).
43. Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression. *Trends in Cell Biology* **19**, 630–637 (2009).
44. Jones, D. & Elf, J. Bursting onto the scene? Exploring stochastic mRNA production in bacteria. *Current Opinion in Microbiology* **45**, 124–130 (2018).
45. Fusco, D. *et al.* Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells. *Current Biology* **13**, 161–167 (2003).
46. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**, 1025–1036 (2005).
47. Katz, N. *et al.* Overcoming the design, build, test (DBT) bottleneck for synthesis of nonrepetitive protein-RNA binding cassettes for RNA applications. *Nature Communications* (2021).

48. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285–298 (2017).
49. Flory, P. J. Introductory lecture. *Faraday Discuss. Chem. Soc.* **57**, 7–18 (1974).
50. Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell* **82**, 2201–2214 (2022).
51. Deniz, A. A. Percolation physics and density transition frameworks converge in biomolecular condensation. *Proceedings of the National Academy of Sciences* **119**, e2210177119 (2022).
52. Shen, Z. *et al.* Biological condensates form percolated networks with molecular motion properties distinctly different from dilute solutions. *eLife* **12**, e81907 (2023).
53. Spann, S., Tereshchenko, M., Mastromarco, G. J., Ihn, S. J. & Lee, H. O. Biomolecular condensates in neurodegeneration and cancer. *Traffic* **20**, 890–911 (2019).
54. Riback, J. A. *et al.* Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028-1040.e19 (2017).
55. Murat, D., Byrne, M. & Komeili, A. Cell Biology of Prokaryotic Organelles. *Cold Spring Harb Perspect Biol* **2**, a000422 (2010).
56. Al-Husini, N., Tomares, D. T., Bitar, O., Childers, W. S. & Schrader, J. M. α -Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies. *Molecular Cell* **71**, 1027-1039.e14 (2018).
57. MacCready, J. S. *et al.* Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* **7**, e39723 (2018).
58. Monterroso, B. *et al.* Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA. *EMBO reports* **20**, e45946 (2019).
59. Wyman, J. & Gill, S. J. Ligand-linked phase changes in a biological system: applications to sickle cell hemoglobin. *Proceedings of the National Academy of Sciences* **77**, 5239–5242 (1980).
60. Ruff, K. M., Dar, F. & Pappu, R. V. Polyphasic linkage and the impact of ligand binding on the regulation of biomolecular condensates. *Biophysics Reviews* **2**, 021302 (2021).

61. Ruff, K. M., Dar, F. & Pappu, R. V. Ligand effects on phase separation of multivalent macromolecules. *Proceedings of the National Academy of Sciences* **118**, e2017184118 (2021).
62. Haynes, C. *et al.* Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLOS Computational Biology* **2**, e100 (2006).
63. Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerging Topics in Life Sciences* **4**, 307–329 (2020).
64. Darling, A. L., Liu, Y., Oldfield, C. J. & Uversky, V. N. Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *PROTEOMICS* **18**, 1700193 (2018).
65. Borchers, W., Bremer, A., Borgia, M. B. & Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Current Opinion in Structural Biology* **67**, 41–50 (2021).
66. La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* **11**, 247–258 (2010).
67. Chew, J. *et al.* C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348**, 1151–1154 (2015).
68. Mankodi, A. *et al.* Myotonic Dystrophy in Transgenic Mice Expressing an Expanded CUG Repeat. *Science* **289**, 1769–1772 (2000).
69. Jin, P. *et al.* RNA-Mediated Neurodegeneration Caused by the Fragile X Premutation rCGG Repeats in *Drosophila*. *Neuron* **39**, 739–747 (2003).
70. Zu, T. *et al.* Non-ATG–initiated translation directed by microsatellite expansions. *Proceedings of the National Academy of Sciences* **108**, 260–265 (2011).
71. Svoboda, P. & Cara, A. Di. Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* **63**, 901–908 (2006).
72. Bevilacqua, P. C. & Blose, J. M. Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annual Review of Physical Chemistry* **59**, 79–103 (2008).

73. Zhou, K., Aertsen, A. & Michiels, C. W. The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiology Reviews* **38**, 119–141 (2014).
74. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA–protein complex. *Nature Structural & Molecular Biology* **15**, 103–105 (2008).
75. Brown, S. D., Fiedler, J. D. & Finn, M. G. Assembly of Hybrid Bacteriophage Q β Virus-like Particles. *Biochemistry* **48**, 11155–11157 (2009).
76. Pumpens, P. *et al.* The True Story and Advantages of RNA Phage Capsids as Nanotools. *Intervirology* **59**, 74–110 (2016).
77. Vasiljeva, I. *et al.* Mosaic Q β coats as a new presentation model. *FEBS Letters* **431**, 7–11 (1998).
78. Vasiljeva, I. The coat of RNA phase Q β as a carrier for foreign epitopes. *RNS fāga Q β apvalks kā svešu epitopu nesējs* (1999).
79. Medina, G., Juárez, K., Valderrama, B. & Soberón-Chávez, G. Mechanism of *Pseudomonas aeruginosa* RhlR Transcriptional Regulation of the rhlAB Promoter. *Journal of Bacteriology* **185**, 5976–5983 (2003).
80. Young, J. W. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc* **7**, 80–88 (2012).
81. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).
82. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).
83. Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology* **151**, 182–195 (2005).
84. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62–66 (1979).
85. Katz, N. *et al.* Synthetic 5' UTRs Can Either Up- or Downregulate Expression upon RNA-Binding Protein Binding. *Cell Systems* **9**, 93-106.e8 (2019).

86. Katz, N. *et al.* An in Vivo Binding Assay for RNA-Binding Proteins Based on Repression of a Reporter Gene. *ACS Synth. Biol.* **7**, 2765–2774 (2018).
87. Bae, W., Yoon, T.-Y. & Jeong, C. Direct evaluation of self-quenching behavior of fluorophores at high concentrations using an evanescent field. *PLOS ONE* **16**, e0247326 (2021).
88. Ogama, T. A beginner's guide to improving image acquisition in fluorescence microscopy. *The Biochemist* **42**, 22–27 (2020).
89. Granik, N., Katz, N., Willinger, O., Goldberg, S. & Amit, R. Formation of synthetic RNA protein granules using engineered phage-coat-protein-RNA complexes. *Nat Commun* **13**, 6811 (2022).
90. Winter, H. H. & Mours, M. Rheology of Polymers Near Liquid-Solid Transitions. in *Neutron Spin Echo Spectroscopy Viscoelasticity Rheology* 165–234 (Springer, Berlin, Heidelberg, 1997).
doi:10.1007/3-540-68449-2_3.
91. Khalil, N., Candia, A. de, Fierro, A., Pica Ciamarra, M. & Coniglio, A. Dynamical arrest: interplay of glass and gel transitions. *Soft Matter* **10**, 4800–4805 (2014).
92. Jawerth, L. *et al.* Protein condensates as aging Maxwell fluids. *Science* **370**, 1317–1323 (2020).
93. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *eLife* **5**, e18413 (2016).
94. Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat Methods* **10**, 119–121 (2013).
95. Klosin, A. *et al.* Phase separation provides a mechanism to reduce noise in cells. *Science* **367**, 464–468 (2020).
96. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
97. Neeli-Venkata, R. *et al.* Robustness of the Process of Nucleoid Exclusion of Protein Aggregates in *Escherichia coli*. *Journal of Bacteriology* **198**, 898–906.
98. Kim, J. *et al.* Fundamental structural units of the *Escherichia coli* nucleoid revealed by atomic force microscopy. *Nucleic Acids Research* **32**, 1982–1992 (2004).

99. Wolf, S. G. *et al.* DNA protection by stress-induced biocrystallization. *Nature* **400**, 83–85 (1999).
100. Janissen, R. *et al.* Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. *Cell* **174**, 1188–1199.e14 (2018).
101. Golmohammadi, R., Fridborg, K., Bundule, M., Valegård, K. & Liljas, L. The crystal structure of bacteriophage Q β at 3.5 Å resolution. *Structure* **4**, 543–554 (1996).
102. Emenecker, R. J., Griffith, D. & Holehouse, A. S. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophysical Journal* **120**, 4312–4319 (2021).
103. Du, F. *et al.* Regulating the T7 RNA polymerase expression in E. coli BL21 (DE3) to provide more host options for recombinant protein production. *Microbial Cell Factories* **20**, 189 (2021).
104. Taylor, N. O., Wei, M.-T., Stone, H. A. & Brangwynne, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophysical Journal* **117**, 1285–1300 (2019).
105. Farag, M., Borchers, W. M., Bremer, A., Mittag, T. & Pappu, R. V. Phase separation of protein mixtures is driven by the interplay of homotypic and heterotypic interactions. *Nat Commun* **14**, 5527 (2023).
106. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annual Review of Biophysics* **37**, 215–246 (2008).

תקציר

בתוך כל תא חי קיימים אברונים החיוניים לצורך ארגון החומר התאי וקיום ריאקציות אנזימתיות שונות בתוך התא. באברונים קלאסיים כמו גרעין התא או גופיף גולג'י, קיימת ממברנה ליפידית (שומנית) שמפרידה אותם מהסביבה התאית ושומרת על תכולתם. בנוסף לאברונים אלה, קיימת קטגוריה נוספת של אברונים חסרי ממברנה. כאן ניתן למצוא אברונים כמו גופיפי-P וגרגירי לחץ (stress granules). אברונים חסרי ממברנה, כפי ששמם רומז, מאופיינים בכך שאין להם ממברנה השומרת אותם נפרדים מהסביבה התאית. למרות זאת, הם עדיין מסוגלים לשמור על הרכבם ועל תפקודם. מחקרים עדכניים בתחום מצאו כי אברונים אלו נוצרים בצורה ספונטנית על ידי תהליך של הפרדת פאזות. בתהליך זה, תמיסה הומוגנית של מולקולות ומלחים נפרדת לפאזות שונות כאשר כל פאזה מאופיינת על ידי ריכוזים שונים של מולקולות ומלחים. נהוג לאפיין תהליכי הפרדת פאזה ביולוגיים על ידי מצבי החומר (מוצק, נוזל וכדומה) של הפאזות הנוצרות. לדוגמה, הפרדת פאזה מסוג נוזל-מוצק משמעותה אברון בעל מאפיינים מוצקים בתוך סביבה נוזלית.

אברונים חסרי ממברנה מסוגים שונים אחראים ישירות, או תורמים לקיומם, של תהליכים תאיים רבים דוגמת שיחבור ושעתוק במגוון צורות חיים. ניתן למצוא גם אברונים כאלה בתהליכי הדבקה על ידי וירוסים ובהיווצרותם של גידולים סרטניים שונים. ההכרה הגוברת בחשיבותם של אברונים חסרי ממברנה הובילה למאות מחקרים שמטרתם לאפיין אברונים טבעיים, ולבנות אברונים סינתטיים. בהקשר של המטרה השנייה, היכולת לבנות אברון חסר ממברנה ולקבוע את תכונותיו תהווה חיזוק משמעותי לתחום הביולוגיה הסינתטית.

במחקר זה פיתחנו אברונים חסרי ממברנה מבוססי מולקולות RNA. תחילה אנחנו מדגימים כי מולקולות RNA סינתטיות המתקפלות למבנה מרחבי המכיל לולאות, מסוגלות לעבור הפרדת פאזה וליצור מבנים דמוי גרגירים (granules), בתוך מבחנה. בעזרת נוקלאוטידים המסומנים פלורסנטית ששולבו ב-RNA, הצלחנו להראות יצירה של גרגירים ולמדוד תכונות שונות אודותיהם כמו גודלם ומספר המולקולות המשוער בתוכם. בנוסף, כאשר מוסיפים למבחנה חלבון המסוגל להתחבר ללולאות ה-RNA באופן סלקטיבי, מצאנו כי חלבונים אלו מתרכזים בגרגירי ה-RNA וכי שילובם מוביל לשינוי בתכונות הגרגירים.

בהמשך, פיתחנו שיטת מדידה ואלגוריתם תואם למטרת מדידה כמותית של הדינמיקה של גרגירי ה-RNA-חלבון שיצרנו. את זאת אנחנו עושים על ידי מעקב ממושך על עוצמת ההארה הפלורסנטית של הגרגירים וללא שימוש בציוד מיקרוסקופיה ייעודי. בחנו את האלגוריתם על אותות מדומים שיצרנו המדמים את ההתנהגות המצופה מהגרגירים הסינתטיים, ובנוסף בחנו אותו על גרגירים שיצרנו בצורה מבוקרת במבחנה וגילינו שהאלגוריתם מסוגל להבדיל בין גרגירים שנוצרו תחת תנאים שונים.

בעזרת שיטה זאת חקרנו בצורה מתודית את הגרגירים הסינתטיים שלנו ומצאנו שתכונותיהם תלויות במספר הלולאות הנמצאות על מולקולות ה-RNA. בנוסף, ביצענו ניסויים המוציאים את הגרגירים מחוץ למצב יציב ומצאנו שהגרגירים מספקים הגנה על המולקולות בתוכם (חלבונים ו-RNA) מפני תהליכים כמו דגרדציה, וכי מולקולות בתוך הגרגירים שורדות זמן רב ביחס למולקולות חופשיות גם בנוכחות חומרים מעכלים.

לבסוף, הוספנו לתוך הגרגירים שלנו רכיב חלבון המכיל אזור חסר-סדר (intrinsically disordered region). חלבונים מסוג זה ידועים ביכולתם ליצור קשרים עם מולקולות רבות ולעבור הפרדת פאזות בצורה ספונטנית לבדם. חקרנו את הדינמיקה ואת התכונות של הגרגירים המכילים את מולקולות ה-RNA שלנו ואת החלבונים הללו ומצאנו את קיומם של שני תהליכי הפרדת פאזה שונים המסוגלים להפריע אחד לשני, או לעזור אחד לשני, כתלות במספר הלולאות הקיימות במולקולות ה-RNA. אנחנו מראים כי על ידי שימוש בתכונה זו, ניתן להגביר את כמות החלבונים המופקים מחיידקי אי-קולי בצורה משמעותית

פרסומים

Granik, Naor, Noa Katz, Or Willinger, Sarah Goldberg, and Roe Amit. "Formation of synthetic RNA Protein granules using engineered phage-coat-protein-RNA complexes". *Nature Communications* (2022) (Related to chapters 1,2,3 of this thesis).

Granik, Naor, Roe Amit. "GelMetrics: An algorithm for analyzing the dynamics of gel-like phase separated condensates". *BMC Bioinformatics* (in review) (Related to chapter 2 of this thesis).

Granik, Naor, Sarah Goldberg, and Roe Amit. "Formation of multiphase RNP granules by intrinsically disordered Q β coat proteins and hairpin-containing RNA". *Nucleic Acid Research* (in review) (Related to chapter 4 of this thesis).

המחקר נעשה במסגרת התוכנית הבין יחידתית למתמטיקה שימושית בהנחיית פרופסור רועי עמית.

אני מודה לקרן גוטווירט ולטכניון – מכון טכנולוגי לישראל על התמיכה הכספית הנדיבה בהשתלמותי.

מחבר/ת חיבור זה מצהיר/ה כי המחקר, כולל איסוף הנתונים, עיבודם והצגתם, התייחסות והשוואה למחקרים קודמים וכו', נעשה כולו בצורה ישרה, כמצופה ממחקר מדעי המבוצע לפי אמות המידה האתיות של העולם האקדמי. כמו כן, הדיווח על המחקר ותוצאותיו בחיבור זה נעשה בצורה ישרה ומלאה, לפי אותן אמות מידה.

אפיון ויישומים של גופי רנ"א-חלבון סינתטיים

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר דוקטור לפילוסופיה

נאור גרניק

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

תמוז תשפ"ד חיפה יולי 2024