

THESIS

CHARACTERIZING UNIQUE FEATURES OF PRION-LIKE DOMAINS RECRUITED
TO STRESS GRANULES

Submitted by

Matthew Henry Baer

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2021

Master's Committee:

Advisor: Eric D. Ross

Timothy J. Stasevich
Mark D. Zabel

Copyright by Matthew Henry Baer 2021

All Rights Reserved

ABSTRACT

CHARACTERIZING UNIQUE FEATURES OF PRION-LIKE DOMAINS RECRUITED TO STRESS GRANULES

Upon nucleation into a cross- β -structure, a classical amyloid will elongate in a concentration-dependent manner as successive hydrogen bonds are made along each exposed β -strand. Solid-state assemblies in biomolecular condensates such as the Balbiani Body in oocytes share a similar β -sheet rich structure that contributes to low dynamicity of the interactions in this separated phase. Prion-like domains (PrLDs) are domains that compositionally resemble yeast prion domains. Some PrLDs can form solid-state assemblies; however, PrLDs have also been associated with the formation of more liquid-like biomolecular condensates. Liquid-like biomolecular condensates form upon varying stimuli, and they appear to assemble through different mechanisms of recruitment than what has been observed for solid-state assemblies. Many solid-state assemblies form via homotypic interactions between identical assembly-prone PrLDs. In contrast, liquid-like assemblies generally form through heterotypic interactions across multiple components. Both solid-state and liquid-like condensates recruit domains that are currently classified as “prion-like”, but evidence is mounting that these contrasting condensates preferentially recruit functionally different types of prion-like domains. A unique feature of prion-like domains recruited to liquid-like condensates is that formation of these assemblies appears to obey a more diverse set of rules for recruitment. This diversity of recruitment of PrLDs may be largely due to the heterotypic interaction

mechanism for assembly into liquid-like condensates. This mechanism of recruitment may also be partially responsible for the different compositional preferences observed for PrLDs recruited to liquid-like assemblies.

PrLD manipulation is a convenient tool for analyzing the compositional elements contributing to formation of both classes of biomolecular condensates. To elucidate the divergent features of PrLDs recruited to liquid-like assemblies, we used systematic mutation of synthetic PrLDs, among other techniques, to examine their unique compositional requirements for assembly. We have created a set of synthetic PrLDs which demonstrates a gradient of assembly-propensity, where localization to stress granules is modulated by differences in residue composition of the domains. To measure the different degrees of assembly propensity of PrLDs, we developed a novel means of quantitative image analysis to determine the enrichment of PrLDs in stress granules. The goals of this work are to help uncover some of the compositional features unique to prion-like domains that are recruited to liquid-like assemblies such as stress granules, to introduce a novel method of image analysis that will allow for quantification of assembly propensity of PrLDs in yeast, and to probe the concentration dependence for assembly of PrLDs into stress granules.

ACKNOWLEDGMENTS

There are many people I would like to thank for helping me during my time at Colorado State University. First, I would like to thank the department of Biochemistry & Molecular Biology for providing an excellent opportunity for undergraduate and graduate education. Thank you to all for showing me that learning biochemistry can be as fun as it is rigorous. I would like to thank Kacy Paul for being an excellent teacher, and a trusted source for assay tweaks and troubleshooting failed experiments. I would like to thank Sean M. Cascarina for being a patient, yet thorough teacher of coding in Python. I cannot thank you enough for teaching me how to code, and how to think about a problem in terms of coding. I would also like to thank my committee members Tim Stasevich, and Mark Zabel for being a part of my student advisory committee. Finally, I would like to thank my advisor Eric D. Ross. Thank you for teaching me how to investigate a scientific question meaningfully, and how to present one's findings as effectively as possible. There is no doubt in my mind that working at The Ross Lab has left a lifelong impact on me, and I appreciate you giving me the opportunity to join all those years ago.

I would also like to thank Anastasia Fomicheva and Miriam Mioro for being great lab mates, and for the fun times in the lab.

Lastly, I would like to thank my wife, Rachel. You are my best friend, and my favorite person to talk biochemistry with. To quote the great Carl Sagan: "In the vastness of space and the immensity of time, it is my joy to share a planet and an epoch with (you)".

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	iv
CHAPTER 1: INTRODUCTION.....	1
Prion-like Domains.....	1
Solid-state assemblies.....	4
Liquid-like condensates and Liquid-Liquid Phase Separation (LLPS).....	10
Concluding Remarks.....	20
CHAPTER 2: SYNTHETIC PRION-LIKE DOMAINS.....	21
Introduction.....	21
Results.....	22
Discussion.....	32
Materials and Methods.....	33
CHAPTER 3: DEVELOPING A NOVEL METHOD FOR AUTOMATED QUANTITATIVE ANALYSIS OF PRION-LIKE DOMAIN LOCALIZATION TO STRESS GRANULES.....	35
Introduction.....	35
YeastSpotter.....	35
Pearson Correlation Coefficient.....	36
Partition Coefficient.....	39
Discussion.....	45
Materials and Methods.....	46
CHAPTER 4: PROBING THE CONCENTRATION DEPENDENCE FOR PRION-LIKE DOMAIN RECRUITMENT TO STRESS GRANULES.....	47
Introduction.....	47
Results.....	47
Discussion.....	57
Materials and Methods.....	60
CHAPTER 5: CONCLUSIONS.....	62
REFERENCES.....	65
APPENDICES.....	69
Running the Pearson Correlation Coefficient method scripts.....	69
Running the Partition Coefficient method scripts.....	76

CHAPTER 1: INTRODUCTION

Prion-like Domains

Devastating neurodegenerative diseases such as Alzheimer's, Parkinson's, or Amyotrophic Lateral Sclerosis share some common features at the cellular level including accumulation of toxic protein aggregates, and the formation of inclusion bodies or amyloids, in subsets of neurons [1]. These amyloids appear to show transmission between cells that is reminiscent of observed characteristics for proteinaceous infectious particles (prions); their aggregation cooccurs with neuronal tissue loss, and amyloid aggregation leading to cell-to-cell transmission is believed to be associated with disease progression observed for these conditions [2].

Insights into the mechanistic basis for these diseases have emerged from studies of prions in yeast. In 1994, the first yeast prions were identified by Reed Wickner. He characterized [*URE3*] and [*PSI*], the prion forms of the Ure2 and Sup35 proteins, respectively [3]. His work illuminated the presence of prions in the proteomes of multiple types of eukaryotic organisms, and not just in mammals as previously observed. The phenotypes resulting from yeast prions had been encountered years earlier with [*PSI*] imparting non Mendelian inheritance and being proposed as a "cytoplasmic mutation" first [4]. Another group later observed ureidosuccinic acid uptake in yeast [5]. These curious phenotypes were recognized by Wickner to be [*PSI*] and [*URE3+*] strains predating the yeast prion classification. This work was followed up to identify the domain of Ure2 contributing to prion aggregation, and led to identification of the "prion-inducing domain" of Ure2 [6].

Researchers investigated Ure2 and Sup35 as models to examine the sequence features that drive prion aggregation. The prion domains of both proteins are enriched in glutamine and asparagine, and depleted in charged and hydrophobic amino acids. Simple bioinformatics methods using these compositional characteristics, combined with phenotypic screens, led to the identification of a third prion, [PIN], which is the prion form of Rnq1 [7]. The compositional similarity between the Ure2, Sup35, and Rnq1 prion domains raised the possibility that amino acid composition could be a primary determinant of prion propensity. Indeed, randomizing sequences in the prion-like domains of Ure2 and Sup35 had no effect on their prion forming ability [8], [9].

Based on these compositional biases, other labs began to develop progressively more sophisticated methods to identify prion-like domains, which are domains with high compositional similarity to known yeast prion domains. Astonishingly, the proteomes of many species including humans and *Saccharomyces Cerevisiae* were found to be equipped with numerous prion-like domains [7], [10].

Yeast prion formation involves the assembly of intrinsically disordered, glutamine and asparagine enriched, low amino acid complexity domains into amyloid fibrils [9]. These residues are particularly important for formation of the hyper-stable dry interface of the amyloid core. When compositional preferences were illuminated as an important feature to seek out when looking for other prion-like domains, common sequence features between human prions and yeast prions began to emerge. These features were discovered by successive rounds of defining criteria among known yeast prion-like domains, followed by discovery of new yeast prions to further define the sequence

features of yeast prion-like domains to seek out in bioinformatic analysis. Remarkably, many of the prion domains that were discovered can efficiently induce prion formation irrespective of primary sequence because their residue composition.

The research community continued to refine their prion aggregation propensity prediction algorithms to reveal compositional requirements for aggregation with greater detail [12], allowing for the identification of hundreds of human proteins with prion-like domains. Mutations located within the prion-like domains many of these proteins, including FUS, TDP-43, hnRNPA1, hnRNPA2, TAF-15, EWSR1, are associated with neurodegenerative diseases [13]. The discovery of many proteins existing at the edge of pathological aggregation led researchers to examine possible physiological implications of aggregation of these domains. Prion-like domains were observed to be recruited to both solid-state assemblies like the Balbiani body [14], or more liquid-like assemblies such as The P Granule [15], and Stress Granules [16]. Although prion-like domains are classically associated with both solid-state and liquid-like assemblies evidence is emerging that suggests that the types of prion-like domains recruited to liquid-like assemblies represents a functionally separate group from the prion-like domains that are recruited to solid-state assemblies. The composition of prion-like domains recruited to liquid-like assemblies results in differences in the mechanism of recruitment to their respective condensate environment, and these mechanistic differences give rise to other features unique to this sub-class of prion-like domains.

Solid-state assemblies

Solid-state assemblies are diverse, however, yeast prions are a highly characterized solid-state assembly that provides insight to general properties. The prevalence of cross- β -structure, or a parallel β -sheet structure, in a majority of yeast prions is a clue that helps explain some of the unique characteristics observed for these hyper-stable assemblies. In-register parallel β -sheet structures result from the interaction of identical prion-like domains with one-another (Figure 1.2). Identical domains bind to the exposed strand in-register so that elongation of the structure proceeds perpendicular to the direction of the β -sheet peptide backbones. Each prion-like domain is added to the parallel β -sheet at an exposed strand, and is incorporated with identical orientation each time (Figure 1.1). This means that each cross- β -structure that forms possesses a unique structure specific to the prion-like domain that contributed to its formation. This allows for many variations in primary sequence of prion-like domains.

It is worth noting that PrLDs represent only a subset of aggregation-prone, intrinsically disordered regions (IDRs), and other IDRs may possess different compositional preferences. However, in order to understand the homotypic interactions that drive recruitment prion-like domains to solid-state assemblies, and why these specific compositional preferences arise from homotypic interactions, Figure 1.1 and Figure 1.2 depict the peptide configuration characteristic of homotypic interactions.

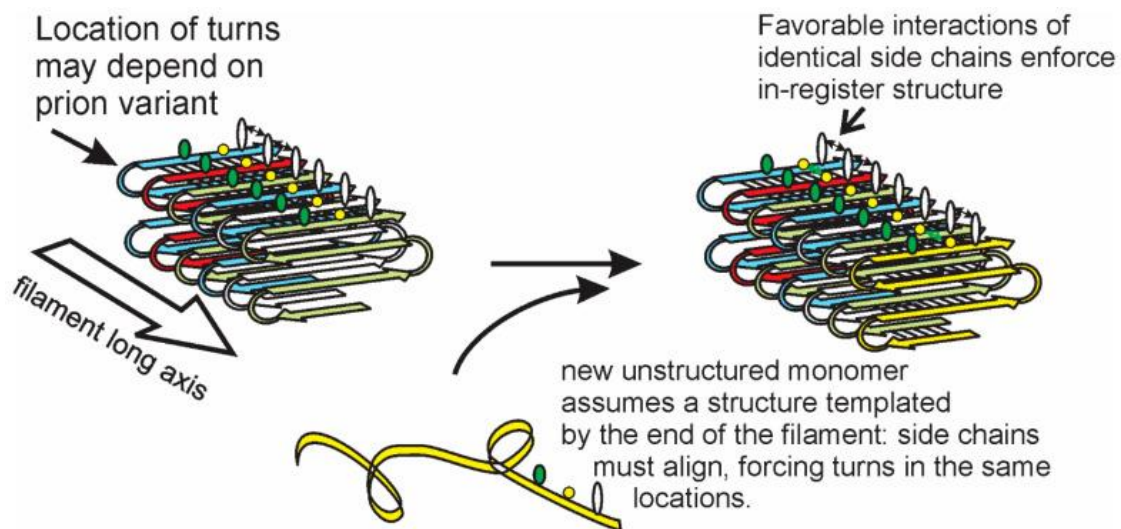


Figure 1.1: Parallel β -sheets can adopt a variety of structures with a common homotypic interaction pattern. Identical domains stack one after another along the filament long axis. From [16].

Parallel β -sheet



β -Bonded partner unchanged by shuffle;
environment may change

Figure 1.2: Parallel β -sheet structures maintain interactions upon scrambling promoting assembly from a variety of amino acid sequences. The maintenance of residue interaction partner specificity is not dependent on primary sequence in these structures, and the interaction pattern resulting from randomizing the residue sequence can be predicted using the following method shown above. From [11].

An in-register parallel β -sheet structure likely explains the insensitivity of yeast prion domains to scrambling (Figure 1.2). Because each amino acid is interacting with its identical counterpart in the adjacent peptide chain, scrambling does not change these primary interactions. Additionally, this illustration allows for inference of how the changes in physicochemical properties introduced by mutation will affect the resulting structure. For example, if a single acidic residue were inserted into a sequence it would be adjacent to its identical acidic counterpart in a neighboring polypeptide chain when attempting to assemble into a parallel β -sheet, and their negatively charged side groups would repel each other in close proximity leading to unfavorable conditions for cross- β -structure formation. This helps to explain the enrichment for uncharged residues in prion-like domains, and explains the commonly observed enrichment for one or more of the following types of residues: branched chain amino acids, and polar residues such as glutamine or asparagine, among other amyloid strain-specific residue features [18]. These residues can form structurally favorable non-covalent interactions in a parallel β -sheet conformation through stabilizing van Der Waals interaction between hydrophobic residues, or hydrogen bonding between uncharged polar residues stacked in a cross- β -structure.

Not surprisingly, there are exceptions to these rules with formation of variant amyloid structures such as amyloidogenic parallel β -helices being observed previously [18], along with the demonstration of anti-parallel β -sheet amyloids that form when aggregation prone domains meet specific amino acid sequence requirements [19], [21]. However, these structures do display either partial β -sheet character, or full β -sheet character, respectively, with stacking of identical domains on one-another similarly to

parallel β -sheet structures. Fibril elongation for parallel β -helices occurs with a helical progression for elongation of the aggregate instead of a linear progression of elongation as is classically observed, but nonetheless fibril elongation occurs with the characteristic domain stacking that is shared among amyloids and solid-state assemblies.

Despite some exceptions in resulting quaternary structure of the solid-state assemblies, the prion-like domains recruited to these assemblies share concentration-dependence for recruitment via homotypic interactions. Solid-state structures that form from homotypic interactions display a strong concentration dependence for structure elongation, where increases in concentration of monomeric amyloid seed, or prion-like domain containing proteins, are highly correlated with increases in amyloid fibril elongation rate [21].

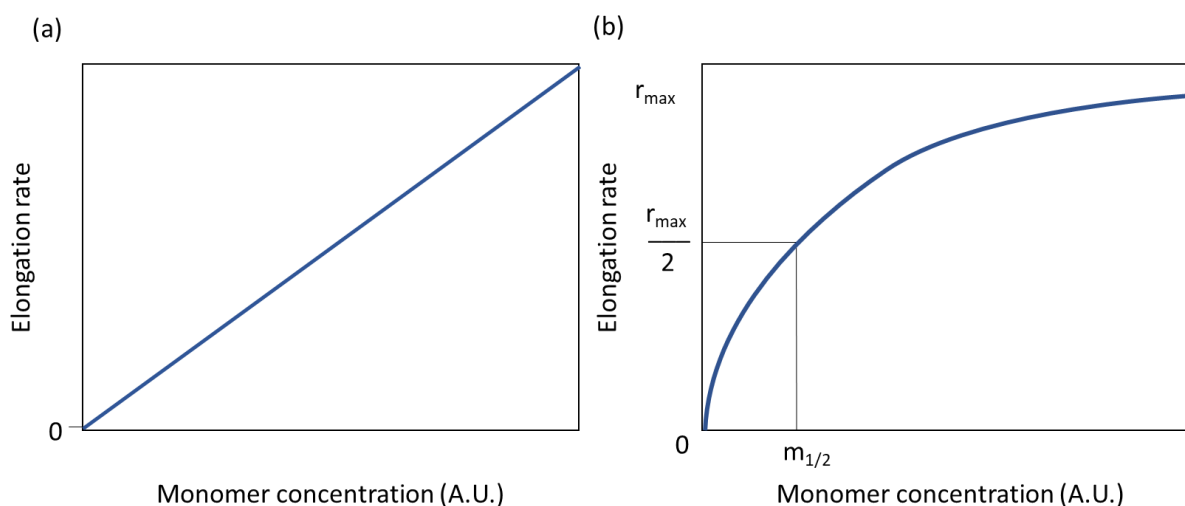


Figure 1.3: Low (a) and High (b) monomer concentration effects on elongation rates of growing amyloid fibrils. Monomer refers to aggregation prone domains (PrLDs) recruited to solid-state assemblies that are characteristic of amyloid fibril formation. The low concentration regime on the left (panel a) shows that at equilibrium monomer concentration, the elongation rate will be zero. Elongation rate displays a linear relationship with concentration at low concentrations of monomer, highlighting the simple concentration dependence for these systems. In panel b, at high concentrations of monomer, the hyperbolic relationship between monomer and elongation rate displays decreases in elongation rate with respect to changes in concentration of monomer during saturation of the sites of active elongation as concentration continues to increase beyond $m_{1/2}$. Figure is not real data and is inspired by a figure from [21].

Importantly, a concentration dependence is observed for the formation of pathogenic amyloids such as the Parkinson's-disease-associated alpha-synuclein [23], among many others. A similar concentration dependence for recruitment to solid-state assemblies such as yeast prions has also been observed by several groups. Likewise, amyloids that facilitate storage of peptide hormones in cross- β -sheets [24], are expected to elongate in a concentration-dependent manner as well. This feature of concentration-dependence for assembly is due in large part to the homotypic interaction mechanism of PrLD recruitment to solid-state assemblies.

As the features of solid-state assemblies began to become clearer, the correlations between domains recruited to either solid or liquid-like assemblies were analyzed. Examination of the interaction mechanisms of these domains leading to their recruitment to different types of condensates was a logical next step. In a 2015 paper from the lab of Dr. Simon Alberti, the prion-like domains of Rnq1 and Lsm4 were found to be recruited to both solid and liquid-like assemblies given the right conditions [25], and this work among other manuscripts largely solidified the view that prion-like domains are recruited to both types of assemblies. However, there is a subtle piece of evidence supporting reclassification in this 2015 manuscript in the parameters needed to induce amyloid formation. The PrLD of Rnq1 formed amyloid only in the presence of active amyloid seed (pre-formed fibrils), and the PrLD of Lsm4 required over expression, or increased monomer concentration, beyond physiologically relevant levels to form amyloid. Under normal conditions for formation of liquid-like compartments such as the P granule, or stress granules, a marked absence of amyloid formation resulting from these reaction conditions was observed. However, it is not clear if persistent stress conditions in the cell lead contribute to pathological aggregation, or whether aggregation is triggered by mutation in normally reversibly-assembling domains.

Thus, it appears that some PrLDs have a distinct preference for solid-state assemblies, while others are preferentially recruited to liquid-like condensates. It is critically important to point out that it is believed that these PrLDs are recruited through multivalent heterotypic interactions with identical domains, and other components such as nucleic acids in the condensate as well. This is opposed to proceeding through concentration-dependent homotypic interactions with only identical domains alone like

what has been observed for classical PrLDs. To better encapsulate the multivalent nature of heterotypic interactions driving recruitment of prion-like domains leading to the formation of liquid-like assemblies, a clear distinction from classical prion-like domains must be demarcated. For one to understand why prion-like domains that are recruited to liquid-like assemblies are functionally unique, it is important to look at some of the current published information on liquid-like assemblies.

Liquid-like condensates and Liquid-Liquid Phase separation (LLPS)

Liquid-like biomolecular condensates represent a functionally separate group from solid-state assemblies. Some contrasting features include differences in structure, and a different mechanism of recruitment (heterotypic, as opposed to homotypic). One striking difference between solid-state assemblies and those that are liquid-like is the on/off rate of monomer interaction with their respective structure. In solid-state assemblies, monomer binds or dissociates at the exposed β -strand. Formation of a cross- β -structure at an exposed β -strand is a very stable interaction, and this leads to a low dissociation constant of monomers from the fibril. In contrast, heterotypic interactions are primarily responsible for driving the formation of liquid-like assemblies through multiple transient contacts between peptides and other condensate components like nucleic acids. Heterotypic interactions bind and dissociate readily, much like water molecules in the liquid phase interacting transiently. This results in a wildly different on/off rate of interaction in heterotypic interactions than the rates observed for homotypic interactions in solid-state assemblies. However, the transient

nature of heterotypic interactions is consistent with the reversibility of stress granule formation, as these structures dissolve when stress is ameliorated.

The first liquid-like condensate characterized and demonstrated to undergo liquid-liquid phase separation was the P granule. The Brangwynne lab showed that P granules undergo droplet fusion, and wetting of surfaces; additionally, they showed that phase separation occurred in a spatiotemporal manner where droplet size corresponded with proximity to cell posterior [15]. Cells have developed means of controlling phase transitions to allow for the formation of organelles by liquid demixing from the solvent they were dissolved in. Liquid-like assemblies display features of liquids such as droplet fusion, a spherical shape due to condensate surface tension [26], and free diffusion with the solvent such as cytoplasm for stress granules, or nucleoplasm for the nucleolus.

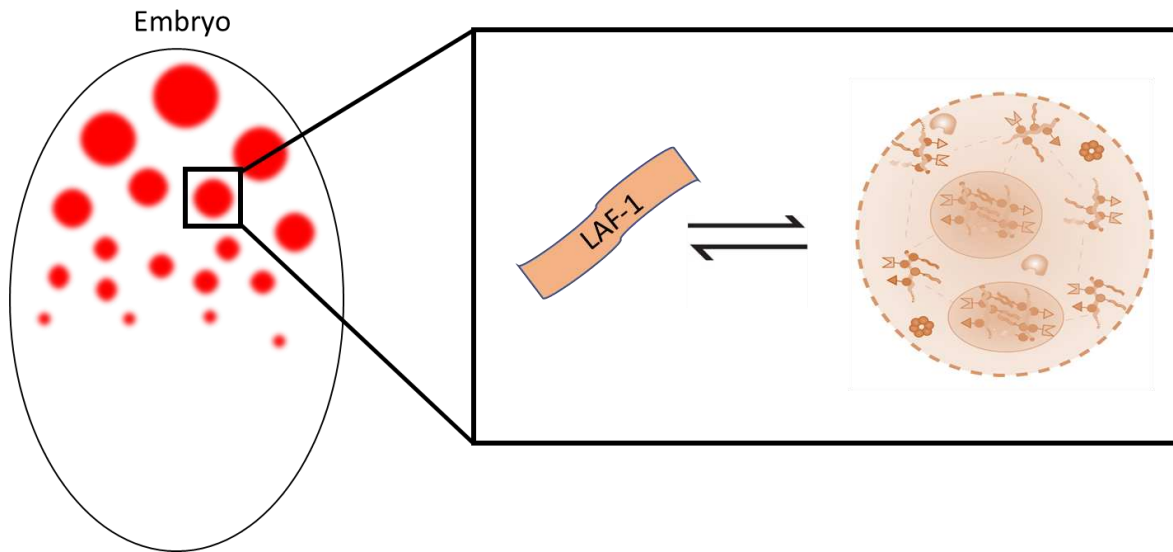


Figure 1.4: P granules form via liquid-liquid phase separation in a spatiotemporal manner. Droplet size is correlated with proximity to cell posterior. LAF-1 is a component recruited through LLPS. Each droplet is comprised of multiple different types of components undergoing heterotypic interactions leading to a dynamic phase equilibrium with the surrounding solvent. The dynamic nature of these interactions results in components readily dissociating from the structure at a higher rate than in solid-state assemblies. Figure inspired by a figure from [26].

Components of liquid-like assemblies are thought to engage in a complex network of multivalent homotypic and heterotypic interactions. Liquid-like assemblies recruit prion-like domains that are often intrinsically disordered, but of a different composition from classical prion-like domains. These prion-like domains have been proposed to engage in diverse heterotypic interactions with other PrLDs, RNA-binding domains, and RNA. The more transient interaction network of a liquid-like assembly allows for more free exchange with the solvent, and heterotypic interactions are responsible for many of the other liquid like properties that differentiate these condensates from solid-state assemblies.

Additionally, it appears that at least in some cases these PrLDs can contain, low-complexity aromatic-rich segments which are believed to be capable of forming homotypic kinked β -sheets. Aromatic residues in PrLDs have been proposed to be participating stabilizing, yet transient, cation- π interactions in kinked β -sheets. Some sequences in protein domains associated with human diseases are also believed to be capable of forming this variant β -sheet structure [27]. However, it is unclear how crucial aromatic residues are for recruitment of PrLDs to stress granules, despite their proposed role in forming these structures. These transient interactions between PrLDs are expected to be only a portion of the total interaction network facilitating recruitment of PrLDs, and thus the primary means of recruitment to liquid-like assemblies is through heterotypic interactions.

Detection of aromatic residues in some intrinsically disordered domains associated with neurodegenerative diseases such as ALS led groups to focus on domains of these type for their potential role in driving assembly of liquid-like structures through liquid-liquid phase separation. The protein FUS was chosen for its ability to aggregate as detected in ALS patients, and Simon Alberti & colleagues proposed that the aggregation propensity of FUS in liquid-like droplets was directly correlated to the valency established by mutation favoring cation- π interactions. They found that mutations introducing polar uncharged residues such as glutamine conferred more solid-like characteristics to the condensates such as lower rates of droplet fusion [28]. These mutations alter the interaction preference of the IDRs from transient interactions to favoring more stable interactions. This modification of composition of the prion-like domains (PLD) in FUS represents a deviation from the usual characteristics of domains

recruited to liquid-like assemblies toward greater polar enrichment, and highlights the role that composition plays in the interaction rates between prion-like domains for both solid and liquid-like assemblies.

The research community now knows that liquid-like assemblies are highly diverse in their heterotypic interaction networks. Even though phase separation can occur with RNA alone [30], often it occurs in the cell along with other proteinaceous components being involved as well. Despite the multicomponent nature of the interactions in liquid-like assemblies formed by phase separation, there are shared compositional features of PrLDs that are recruited to stress granules. These sequence features can be used to predict whether a domain will be recruited to stress granules based only on the amino acid composition.

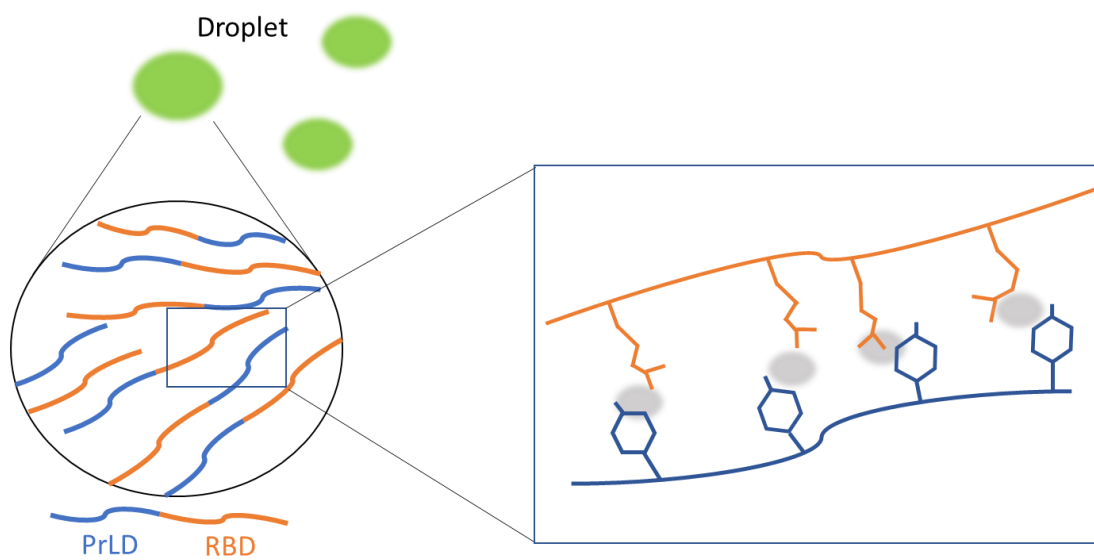


Figure 1.5: PrLD-RBD: Prion like domain-RNA binding domain fusions. Domains are undergoing LLPS to form separate phase droplets. Cation-pi interaction networks are varied depending on residue composition. Cation-pi interactions between tyrosine and arginine are believed to be important for LLPS. Figure inspired by a figure from [28].

Recent generation of a dataset comprised of yeast PrLDs recruited to stress granules upon cell stress, as well as domains which remained diffuse, illuminated the compositional preferences of PrLDs recruited to stress granules ([29]; Table 1.1). This manuscript also described the construction of synthetic prion-like domains (sPrLDs) recruited to *bona fide* stress granules to verify the observed compositional preferences. In addition to synthetic assembling PrLDs, non-assembling control PrLDs (cPrLDs) were constructed to verify that a composition reflecting non-assembling PrLDs would hinder recruitment to stress granules. Indeed, cPrLDs remained diffuse after formation of

stress granules in cells. For the sPrLDs, colocalization with a known stress granule marker confirmed sPrLD recruitment. An absence of assembly for control domains verified that colocalization with the stress granule marker was dependent on a composition reflecting assembling PrLDs. Compositional changes to these synthetic domains are predictive of changes in phenotype, suggesting a change in interaction preferences of the prion-like domains. These changes in composition can either favor or hinder participation in heterotypic interactions to affect recruitment of PrLDs. Interestingly, aromatic residues were not required for assembly of these synthetic domains, although it was previously believed that aromatic residues were critically important for recruitment to stress granules. Aliphatic residues such as leucine, valine, or isoleucine could functionally replace aromatic residues in supporting recruitment [27]. This is significant because the synthetic domains are recruited to stress granules, a liquid-like assembly. Additionally, these residues had not been previously predicted to support recruitment of PrLDs to stress granules in yeast.

Table 1.1: Compositional preferences for yeast PrLDs tested for colocalization with a known stress granule marker (Pab1). Assembling PrLDs are classified by colocalization with Pab1 after stress. Non-assembling PrLDs are classified by remaining diffuse after introduction of cell stress. The mean frequency of each residue is calculated as the mean frequency of appearance in each type of domain (assembling, or non-assembling) within the dataset. Residues in orange are more overrepresented in non-assembling domains. Residues in blue are more overrepresented in domains that assemble after introduction of cell stress. Adapted from [31].

Amino Acid	Mean frequency in assembling PrLDs	Mean frequency in non-assembling PrLDs
A	2.79	5.38
C	1.22	0.23
D	4.58	4.33
E	4.11	2.44
F	4.71	2.41
G	5.27	6.63
H	2.94	2.06
I	5.23	2.78
K	4.83	3.12
L	6.07	4.26
M	1.67	2.98
N	13.06	16.15
P	3.94	8.30
Q	6.79	13.91
R	6.04	3.07
S	12.57	9.57
T	4.46	4.54
V	5.19	2.72
W	1.22	0.30
Y	3.31	4.83

One of the most striking compositional deviations for PrLDs observed in Table 1.1 is the preference for the residues isoleucine, leucine, and valine in assembling domains. In addition to this, phenylalanine and tryptophan, as well as charged residues are over-represented in assembly-prone PrLDs [31]. PrLDs recruited through heterotypic interactions to stress granules appear to display a preference for charged residues, a preference not observed in PrLDs that are associated with yeast prion formation. This is a significant deviation from a classical prion-domain, as charged residues are not conducive to homotypic interactions in cross- β -structures because of identical side-chain repulsion. Prion-like domains recruited through heterotypic interactions not only differ in mechanism of recruitment, but also show clear compositional deviations from domains recruited to solid-state assemblies through homotypic interactions. These differences suggest that there are distinct classes of prion-like domains; since our current methods of identifying prion-like domains are derived from studies of yeast prions, they may be poorly suited for identifying prion-like domains recruited to liquid-like assemblies through heterotypic interactions.

Another recent manuscript by the lab of Clifford Brangwynne demonstrated heterotypic interactions leading to liquid-liquid phase separation of the nucleolar component NPM1 [32]. Their work showed that liquid-like assemblies follow a different set of guidelines governing assembly propensity; mainly that recruitment to liquid-like assemblies is dependent on the concentrations of multiple components, or the effective concentration of interacting partners, and not a simple dependence on concentration of monomer observed for homotypic interactions. Heterotypic interactions are concentration dependent, but in a different manner. The interactions are believed to be

interchangeable among numerous components, and as a result the individual domains may possess a functionally static concentration of interacting partners due to the interchangeability of interacting partners as well as their abundance. This is in stark contrast to the simple concentration-dependence observed formation of solid-state assemblies through heterotypic interactions.

Therefore, in this thesis we examined the concentration dependence of PrLD recruitment to stress granules, as well as the compositional features that support recruitment. Our preliminary data supports the hypothesis that recruitment to liquid-like assemblies through heterotypic interactions does not display a simple concentration-dependence of prion-like domains *in vivo*. Changes in concentration of prion-like domains have a modest effect on the formation of stress granules. Taken together with the data presented from the Brangwynne lab pertaining to the nucleolus, these results support the hypothesis that recruitment to liquid-like assemblies can be carried out at a variety of concentrations of prion-like domain within the cell. The rules governing recruitment of prion-like domains to liquid-like assemblies through heterotypic interactions are diverse, and still being defined. However, it has become clear that the prion-like domains recruited to liquid-like assemblies are functionally unique from classical prion-like domains that are recruited to solid-state assemblies. These differences highlight a clear separation as a functionally distinct group of domains separate from classical PrLDs.

Concluding Remarks

The chapters following this introduction will describe the interrogation of the necessary and sufficient aliphaticity for recruitment of synthetic PrLDs to stress granules, the development of a novel means of image analysis for quantification of colocalization with a stress granule marker, and the probing of concentration-dependence for assembly of PrLDs by varying expression levels in cells. In chapter 2, we describe the generation of synthetic PrLDs, and the generation of a set of sPrLDs that display an assembly propensity gradient. The assembly propensity gradient was created through serial deletion of aliphatic residues in sPrLDs. In chapter 3 we then describe the generation of a novel, automated means of quantitative image analysis for analyzing subtle changes in colocalization with a known stress granule marker. This image analysis technique then allowed us to meaningfully probe the concentration-dependence for assembly with a more accurate measure of partitioning in chapter 4. Finally, in chapter 5 we conclude with what was learned from these experiments, the implications these findings may have for the study of liquid-like biomolecular condensates, and future directions for this work.

CHAPTER 2: SYNTHETIC PRION-LIKE DOMAINS

Introduction

In our previous screen for PrLDs that are recruited to stress granules, we observed statistically significant compositional differences between PrLDs that are recruited to stress granules and those that are not (Table 1.1; [29]). To test whether these compositional biases are sufficient to determine recruitment propensity, our lab constructed a set of synthetic PrLDs, or control PrLDs designed to approximately match the average compositions of naturally-observed stress-granule-recruited or non-recruited PrLDs, respectively. Each domain is 100 residues in length and the primary sequence was randomized. Andrew Tonsager and Kacy Paul constructed these domains, and performed preliminary microscopy experiments to confirm expression. I completed these microscopy experiments and used western blot to verify domain expression. Once expression of the synthetic and control domains was verified, I tested these domains for recruitment to stress granules. To test the hypothesis that aliphatic and aromatic amino acids were key drivers of PrLD recruitment to stress granules, we created aliphatic and aromatic deletion mutants. The aliphatic residues isoleucine, leucine, and valine were deleted to test the contribution of these residues toward recruitment. Additionally, all three aromatic residues (phenylalanine, tryptophan, and tyrosine) were deleted from sPrLDs to test the contribution of these residues to recruitment of sPrLDs.

Results

In order to explore the compositional preferences of yeast PrLDs and the degrees of compositional enrichment that are necessary to facilitate robust recruitment or lack of recruitment to stress granules, domains with compositions reflecting the mean frequency of each amino acid in stress-granule-recruited or non-recruited PrLDs (Table 1.1) were designed as a “baseline” composition starting point. This resulted in three “mean frequency” assembly-prone PrLDs that were predicted to be recruited to stress granules, and one “mean frequency” non-assembling PrLD which that was predicted to remain diffuse upon heat shock. In addition to these “mean frequency” domains, domains slightly or substantially enriched in the residues overrepresented for their desired phenotype were also constructed to increase the likelihood of a desirable domain behavior. These domains were denoted as “+” or “++” for the assembling domains that were enriched either a little, or slightly more, respectively, toward a composition of residues overrepresented in assembling domains. Enrichment for residues overrepresented in non-assembling domains was also performed for the cPrLDs as well. These were denoted as “-” and “--” to reflect domains that were enriched either a little, or more substantially, respectively, in amino acids overrepresented in non-assembling domains. Each PrLD was fused to the C-terminus of GFP and transformed into *S. Cerevisiae* lab strain (YER1405), in which the stress granule protein Pab1 is tagged with mCherry (See Materials & Methods). Each of the predicted assembly-prone PrLDs showed no detectable expression by western blot, and a very weak signal by microscopy that was consistent with autofluorescence. This lack of clear expression of sPrLDs in the initial phenotype testing led to the design of two

more sPrLDs, and their compositionally-enriched counterparts, to see if different randomized sequences could allow for more clear and detectable expression in cells.

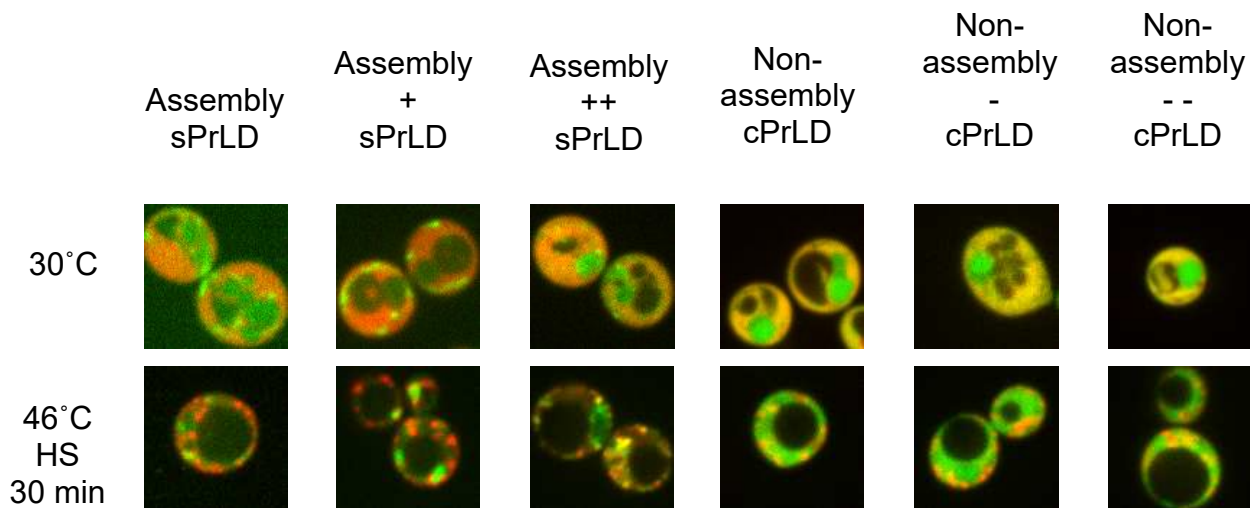


Figure 2.1: “Mean frequency” domains (Assembly sPrLD & Non-assembly cPrLD) displayed with the slight enrichment domains “+” & “-” as well as the further enriched “+ +”, and the “- -” domains. Expression of GFP-tagged sPrLDs was too low to rule out autofluorescence while imaging the assembling domains. sPrLDs, cPrLDs were tagged with GFP, Pab1 was tagged with mCherry. Images are a merge of both channels.
HS: Heat shock

From the expanded set of synthetic assembling domains, three domains (Assembly 2, Assembly 2++, and Assembly 3++) showed a weak signal by microscopy that was consistent with autofluorescence and little or no detectable expression by western blot;

three others (Assembly 3, Assembly 2+, and Assembly 3+) showed a diffuse signal prior to heat shock and a weak but detectable band by western blot (Figure 2.2).

Thus, a total of three predicted non-assembling PrLDs (non-assembly 1, 2 and 3) and three predicted assembly-prone PrLDs (Assembly 3, Assembly 2+, and Assembly 3+) showed detectable expression. All six were initially diffuse upon expression in yeast in unstressed conditions. Upon heat shock, each of the non-assembling PrLDs remained diffuse (Figure 2.1), while each of the assembly-prone sPrLDs assembled into foci that co-localized with Pab1 (Figure 2.2). Since only one of the three constructs in the initial assembly-prone set and none in the ++ set showed detectable expression, for subsequent analysis, we focused on the “+” dataset (Assembly 2+ and Assembly 3+). The domains were subsequently relieved of their cumbersome, lengthy names and dubbed sPrLD2 & sPrLD3. Their identical compositions serve as perfect platform for interrogation of compositional features necessary for recruitment.

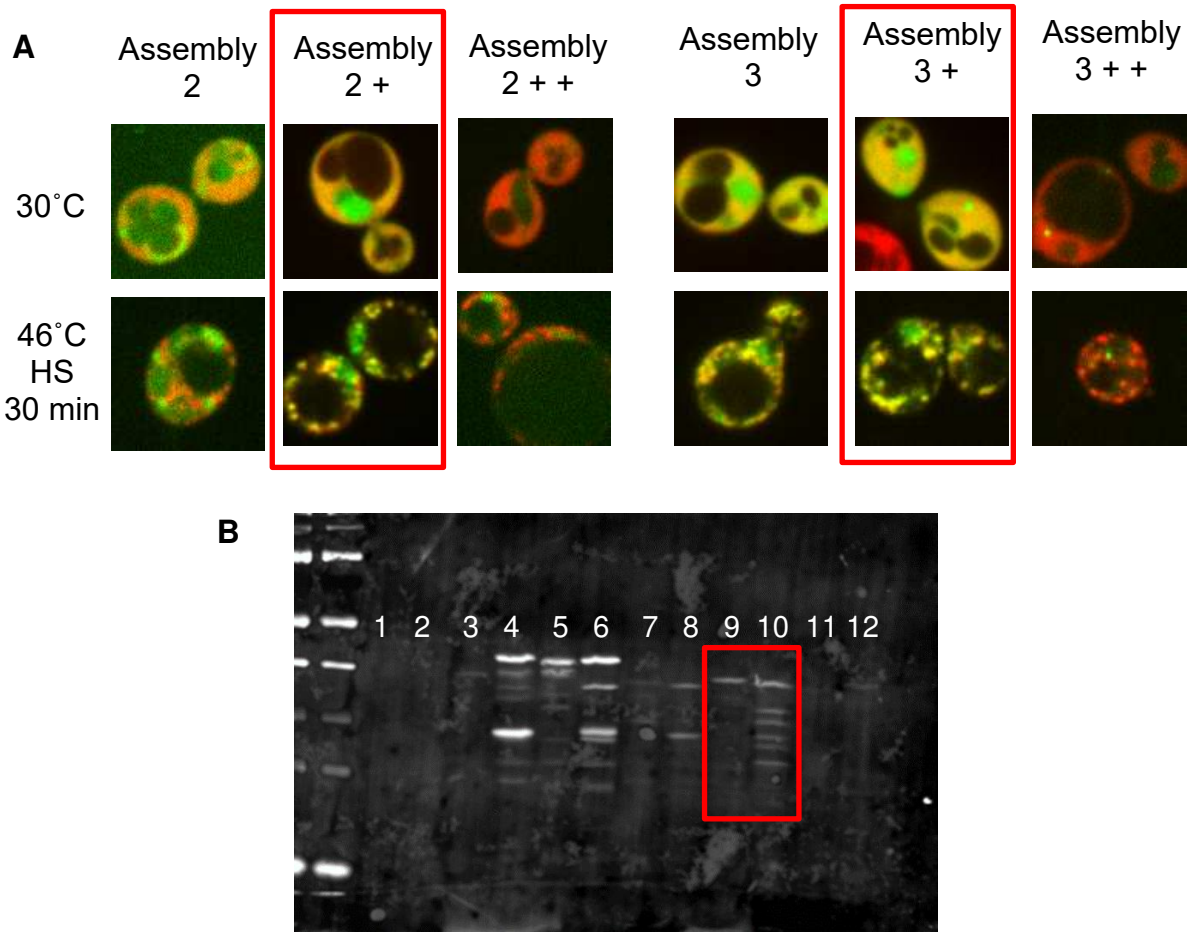


Figure 2.2: (A) Two more “baseline” mean frequency assembling synthetic PrLDs were created (Assembly 2,3), along with their slightly enriched (2+,3+), and severely enriched (2++,3++) counterparts. The slightly enriched domains highlighted in the red boxes were chosen as a set of two identical domains to use for mutation analysis. Colocalization of the highlighted sPrLDs with a known stress granule marker indicates recruitment of sPrLDs to stress granules. Assembly 2+ was renamed to sPrLD2, and Assembly 3+ was renamed to sPrLD3. sPrLDs were labeled with GFP, and Pab1 was labeled with mCherry. **(B)** Western blot of multiple synthetic domains tagged with GFP. Lanes highlighted in the red box indicate Assembly 2+ (sPrLD2) and Assembly 3+ (sPrLD3), respectively. **Lane contents:** 1: Assembly 1, 2: Assembly 1+, 3: Assembly 1++, 4: Non-assembly, 5: Non-assembly -, 6: Non-assembly - -, 7: Assembly 2, 8: Assembly 3, 9: Assembly 2+ (sPrLD2), 10: Assembly 3+ (sPrLD3), 11: Assembly 2++, 12: Assembly 3++. Western blot was carried out as described in Materials & Methods.

Once we had a validated set of two synthetic assembling domains (sPrLD2 & sPrLD3) to go along with a control non assembling domain, we began to ask what compositional features are necessary and sufficient for recruitment of sPrLDs to stress granules. At that time, aromatic residues were believed by many to be drivers of phase separation leading to recruitment to liquid-like assemblies such as stress granules [28], [29]. This informed our initial curiosity to delete aromatic residues, and see if recruitment could still be supported by sPrLDs. In addition to deleting aromatic residues, we also deleted the aliphatic residues isoleucine, leucine, and valine because of the degree to which those residues were overrepresented in natural PrLDs that assembled upon stress (Table 1.1; [31]). To round out the deletion set, we also deleted both the aromatic residues, and the aliphatic residues listed previously. Each deletion mutant was then tested by microscopy for colocalization with a stress granule marker (Figure 2.3 A).

Deletion of either aromatic or aliphatic amino acids, or deletion of both blocked stress granule recruitment, demonstrating that both of these types of residues promoted recruitment of sPrLDs to stress granules. Because both types of residues were enriched in stress-granule recruited PrLDs (Table 1.1), we wondered if aliphatic residues could replace aromatic residues in supporting stress granule recruitment. Therefore, we generated constructs in which all of the aromatic amino acids were replaced with I, L, or V. Building from these initial results in (Figure 2.3 A), we also asked whether any of the aromatic residues were particularly supportive of PrLD recruitment to stress granules. Stress granules contain both protein and RNA [28]. In vitro, RNA promotes phase separation, and tyrosine residues are believed to be particularly stabilizing for phase separation. Interactions with the aromatic rings in RNA leading to recruitment are

believed to proceed at least partially through transient pi-pi interactions with tyrosine residues in PrLDs [27]. However, tyrosine was the only aromatic amino acid not significantly enriched among our stress-granule recruited PrLDs. Therefore, to examine the relative ability of each aromatic amino acid to promote stress granule recruitment, we replaced all of the aromatic amino acids with either Y, F, or W. Finally, we also substituted the aromatic residues for glutamine and asparagine as these residues were overrepresented in non-assembling yeast PrLDs (Table 1.1).

From these substitution mutants, we observed robust recruitment for all of the aromatic-enriched sPrLDs, as well similarly robust recruitment of the aliphatic-enriched aromatic substitutions (Figure 2.3 B). We were surprised to see replacement of aromatic residues with I, L, and V still supported recruitment to stress granules to a similar degree. This was a curious result given the established importance of tyrosine and aromatic residues in general for phase separation as described in the literature at that time.

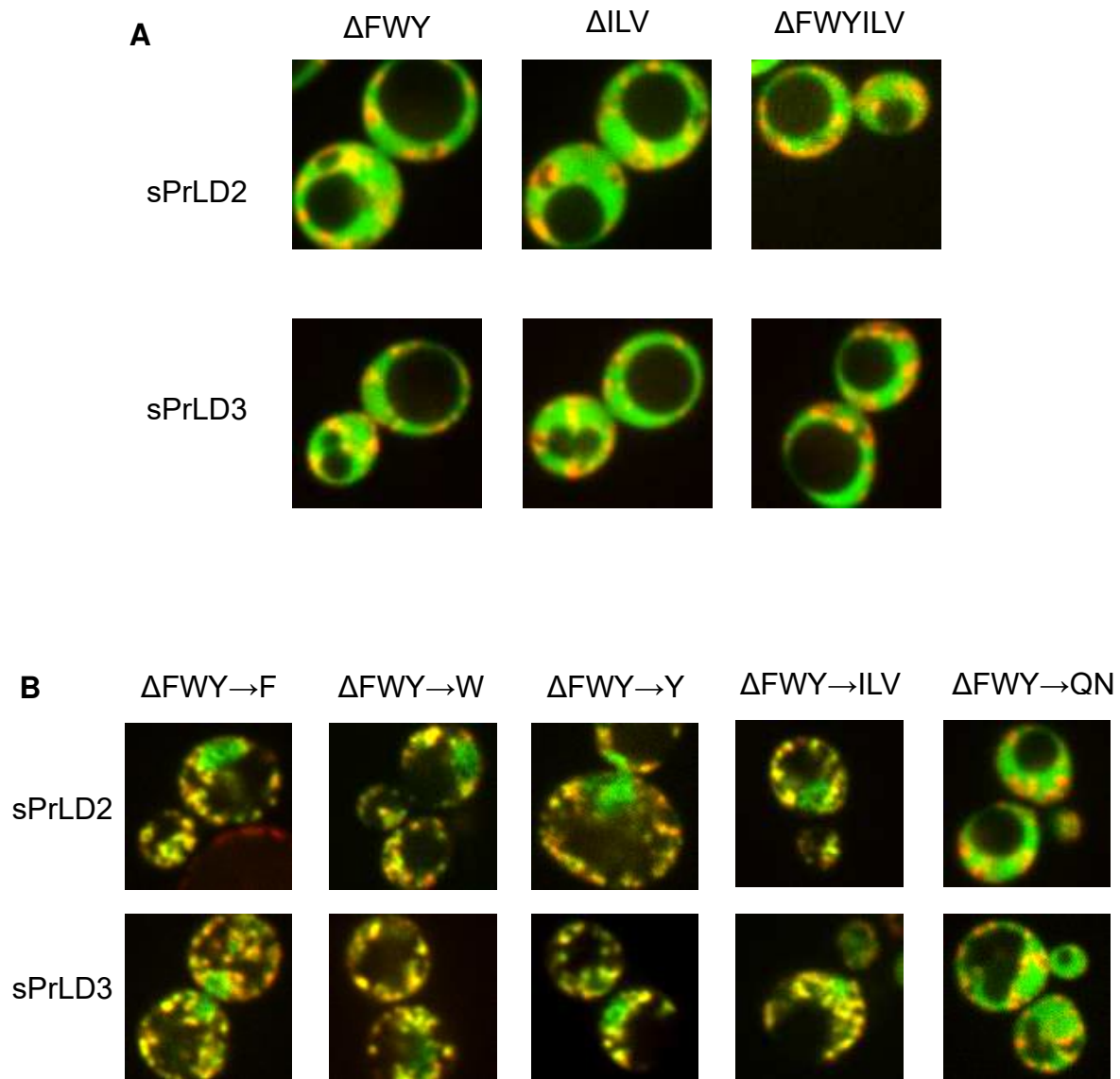


Figure 2.3: (A) Deletion of aromatic residues (Δ FWY) abolishes recruitment of both synthetic domains to stress granules after heat shock. Similarly, deletion of isoleucine, leucine, and valine also disrupts recruitment to stress granules for both sPrLDs. Deletion of both aromatic and select aliphatic residues also disrupts recruitment to stress granules. **(B)** Enrichment for any one aromatic residue seems to have little effect on recruitment of sPrLDs to stress granules. Δ FWY \rightarrow ILV domains displays functional replacement of aromatic residues with aliphatic residues in supporting recruitment to stress granules for both sPrLDs.

We next examined the threshold for aliphaticity required to support stress granule recruitment. The initial sPrLDs contain (19% I, L, V). To determine whether we could titrate the efficiency of recruitment, we serially deleted aliphatic amino acids in sets of 3. These constructs were designated “– 3 ILV”, “– 6 ILV”, ... “– 15 ILV”, and finally “– 19 ILV” (the previously described construct in which all of the I, L, and V are deleted). To design the deletions, three I, L, or V residues were randomly selected for deletion; because all three showed similar levels of enrichment among the stress-granule-recruited PrLDs, each residue type was given an equal probability of being chosen. This led to random triplets of I, L, or V being chosen for deletion, and a slightly different aliphatic profile for each domain due to random selection of I, L, or V residues at equal probability (Figure 2.4 A & B).

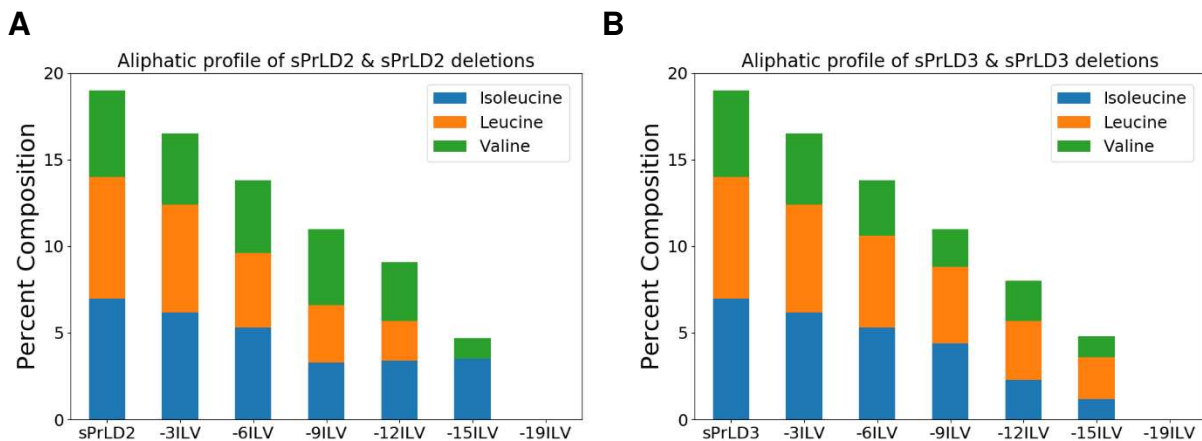


Figure 2.4: (A) Aliphatic profile of sPrLD2 deletions with equal probability of three randomly chosen I, L, or V residues being selected for deletion from the domain each time. **(B)** Aliphatic profile of sPrLD3 deletions following the same deletion pattern, but displaying a different ratio of I:L:V from sPrLD2. Despite ratio differences, both deletion series maintain equal percent composition of selected aliphatic residues.

When the sPrLD 2 & 3 deletion series was imaged by confocal microscopy, the outcome was phenotypically just as predicted (Figure 2.5). We saw a gradual loss of recruitment for both sPrLDs as the domains became less aliphatic through deletion. This result demonstrated tunable colocalization with a known stress granule marker through changes in the number of aliphatic residues leucine, isoleucine, and valine in sPrLDs. These residues were not previously characterized as providing a significant contribution toward recruitment to stress granules. Thus, we sought to measure the subtle changes in colocalization we had produced through deletion of these aliphatic residues. While the gradient of assembly propensity was visible by eye, to accurately convey the subtle changes in assembly we decided that we needed to develop an automated system for image analysis that could quantify the degree of colocalization of sPrLDs with Pab1 in cells. Pursuit of a novel image analysis technique was quite an easy decision as a tool of this kind would prove to be incredibly useful for PrLD colocalization analysis, and we simply could not pass up the potential opportunity to have a tool like this at our disposal.

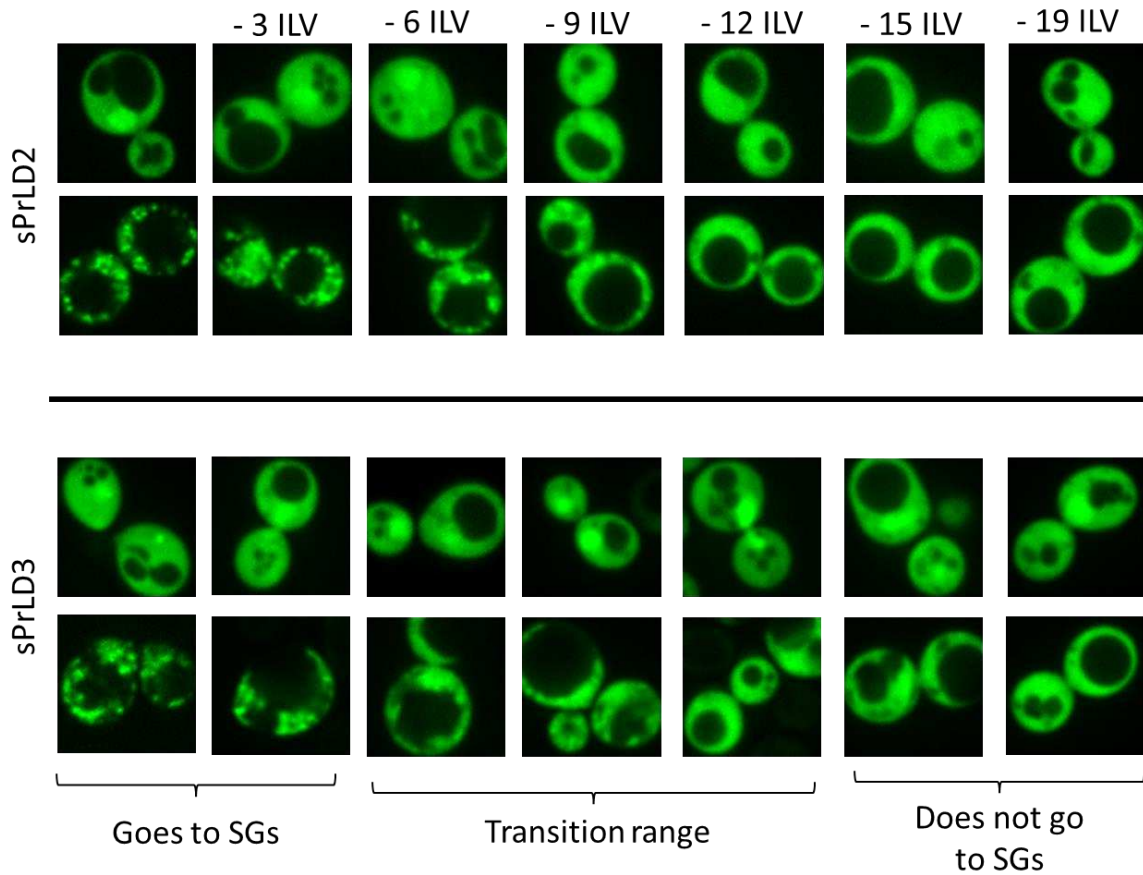


Figure 2.5: Deletion of the aliphatic residues isoleucine, leucine, and valine from sPrLD2 and sPrLD3 results in a gradual loss of assembly into stress granules. For each sPrLD, the top row of images displays the pre-stress phenotype (30°C), and each bottom row of images displays the cells after stress conditions (46°C for 30min). Clear granule formation is observed in the images labeled “Goes to SGs”, and PrLDs remain diffuse after stress in the images labeled “Does not go to SGs”. Images in the labeled “Transition range” display the most subtle differences in assembly, and require careful measurement of colocalization for accurate depiction of the subtle differences in assembly propensity of each domain in the “Transition range”. Images shown in this figure display only the GFP channel for clarity.

Discussion

Our initial excitement over functional replacement of aromatic residues with selected aliphatic residues in synthetic PrLDs [31] led to the decision to focus on these select aliphatic residues for their contribution to recruitment to stress granules. With each residue displaying a roughly similar overrepresentation among yeast PrLDs that assemble upon stress in Table 1.1, we selected three random I, L, or V residues at a time to delete from each domain. These deletions led to the generation of an assembly propensity gradient for both sPrLDs. To our knowledge, aliphatic hydrophobicity had not been significantly associated with stress granule recruitment in yeast. Thus, we were thrilled to be able to begin to characterize the necessary and sufficient aliphaticity for stress granule recruitment through mutation and testing with microscopy. However, these claims were lacking quantitative evidence for support, and in chapter 3 we set out to create a novel means of automated image analysis to robustly measure the assembly propensity gradients we had produced in chapter 2. A tool that would allow for accurate measurement of the assembly propensity gradients is essential to begin to characterize the subtle differences in assembly of sPrLDs in this dataset. As we will see later, a means of quantitative analysis for measuring assembly propensity will also open the door for even more exciting questions we will address in chapter 4.

Materials and Methods¹

Strains & Growth Conditions

The experiments discussed in this chapter were carried out using the model organism *Saccharomyces cerevisiae* grown in standard yeast media, and with standard growth conditions.[33] Plasmids were transformed into the yeast strain YER1405 (MATa his3D1 leu2D0 met15D0 ura3D0 PAB1-mCherry::URA3) using the standard techniques detailed in [33]. YER1405 is a derivative of BY4741 [32]. Yeast were grown at 30°C in pre-stress conditions, or heat shocked at 46°C for 30 minutes before imaging.

Plasmid construction

Generation of plasmid backbone for sPrLD expression (pER843) was carried out by inserting GFP between the HindIII and BamHI restriction sites following the *SUP35* promoter in the plasmid pJ526 from [33]. sPrLDs were then cloned into pER843 in frame with GFP at its N-terminus between the BamHI and BglII sites.

Designing sPrLDs & cPrLDs

Each sPrLD and cPrLD is 100 amino acids long. The average frequency of each amino acid among the assembling and nonassembling PrLDs (Table 1.7) were used as a starting point for design of the sPrLDs and cPrLDs, respectively. To achieve whole-number values for each amino acid, frequencies were rounded in the direction predicted to promote or inhibit assembly for the sPrLDs and cPrLDs, respectively; the numbers were then modestly adjusted such that the final values added up to 100. An Excel random number generator was used to randomly order the amino acids. Sequences were built synthetically using overlapping primers and then cloned into pER843 as described above.

Stress Conditions

Cells were grown to midlog phase at 30 °C in SC-Leu media (to select for pER843) prior to stress induction. For heat shock, 1 mL of cells was concentrated to 50 µL and incubated in a 46 °C water bath for 30 min prior to imaging.

¹This section is reformatted from the following publication: Boncella A.E., Shattuck J.E., Cascarina S.M., Paul K.R., Baer M.H., Fomicheva A, Lamb A.K, Ross E.D. (2020). Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proceedings of the National Academy of Sciences* 117 (11) 5826-5835. My contribution to that work was building and testing of the following constructs under heat shock: Δ FWY, Δ FWY→ILV, Δ FWY→QN. As well as imaging sPrLD2, sPrLD3, cPrLD1 under heat shock. Aliphatic deletion domains were cloned into pER843 using the same method listed in this section.

Confocal Fluorescence Microscopy

Imaging was performed on an Olympus (IX83) Inverted Spinning Disk Confocal Microscope using a 100× objective, and a 488nm laser along with a 561nm laser. All images were captured as single planes.

Western Blotting

Cells were grown to midlog phase at 30 °C in SC-Leu media (to select for pER843) before harvesting. Volumes collected for each culture were normalized to the lowest density culture, as assessed by OD600 measurements. Cells were harvested by centrifugation at 3,000 rpm for 5 min at 4 °C. Cell lysis was performed using a SUMEB lysis buffer. A volume of 15 µL of each sample was run on a polyacrylamide gel and then transferred onto a PVDF membrane. GFP-PrLD fusions were probed using an anti-GFP antibody (Santa Cruz Biotechnology). Images were acquired using Li-Cor Odyssey Clx imaging system.

²This section is reformatted from the following publication: Boncella A.E., Shattuck J.E., Cascarina S.M., Paul K.R., Baer M.H., Fomicheva A, Lamb A.K, Ross E.D. (2020). Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proceedings of the National Academy of Sciences* 117 (11) 5826-5835. My contribution to that work was building and testing of the following constructs after heat shock: Δ FWY, Δ FWY \rightarrow ILV, Δ FWY \rightarrow QN. As well as imaging sPrLD2, sPrLD3, cPrLD1 after heat shock. Aliphatic deletion domains were cloned into pER843 using the same method listed in this section.

CHAPTER 3: DEVELOPING A NOVEL METHOD FOR AUTOMATED QUANTITATIVE ANALYSIS OF PRION-LIKE DOMAIN LOCALIZATION TO STRESS GRANULES

Introduction

In chapter 2 we demonstrated that it was possible to generate a gradient of assembly propensity gradient through serial deletion of isoleucine, leucine, and valine residues in sPrLDs. However, although subtle changes in colocalization were visible to the eye in images, we thought there must be a way to devise an automated means for quantitatively measuring subtle changes in colocalization with a stress granule marker. This required the generation of an automated method for image analysis. To accurately quantify colocalization between PrLDs and a stress granule marker, we needed to be able to first isolate the regions of images that contained cells. Once we could isolate pixels that contained cells, we could determine the degree of colocalization on a cell-to-cell basis in sets of images.

YeastSpotter

Quantification of the colocalization between fluorophores in a microscopy image first requires one to be able to identify the pixels that correspond to the cell one wishes to analyze. In a microscopy image of cells, there are two groups of pixels that contain two different types of information. There are pixels that make up the actual image of the cells, and the rest of the pixels make up the “background” of the image, or the area that does not contain any cells. To accurately isolate yeast cells in microscopy images, we utilized YeastSpotter, an image segmentation tool that uses a trained neural network to identify yeast cells in images [34]. This tool creates masks that can be applied to images to “delete” the background pixel information. With this tool, we only added a

small modification to our workflow to minimize changing our experimental set-up. Each DIC image is captured using collimated white light at low intensity relative to the monochromatic lasers. This additional light exposure is expected to impart low phototoxicity to cells to minimize confounding of results. The image masks created by YeastSpotter from the DIC images are then used for isolating the “cell” pixels in the corresponding fluorescence images used for quantitative analysis.

Once we had established that we could accurately segment images of yeast cells using YeastSpotter, we were able to begin to ask how we would quantify each set of pixels corresponding to the individual cells we were interested in analyzing. We started with a well established method of quantitative colocalization analysis: The Pearson Correlation Coefficient.

Pearson Correlation Coefficient

One of the earliest publications describing a quantitative measurement of colocalization between fluorophores in microscopy images was by Manders et al in 1992 [35]. They plotted the pixel intensity values for each color channel for a given image and found the Pearson Correlation Coefficient between the plotted pixel intensities. This method is useful for colocalization analysis because it describes the degree of overlap between two color channels, irrespective of average intensity differences between channels. Using the Pearson Correlation Coefficient as a preliminary scoring metric for our microscopy data, we obtained promising initial quantification data. The median Pearson Correlation Coefficients decreased as assembly propensity of PrLDs decreased. This confirmed that we could quantify changes in assembly propensity for the sPrLD deletion mutants (Figure 3.1).

This scoring metric was useful in initial exploratory efforts to define the trend we should expect when applying other scoring metrics to our microscopy data. However, we did not end up using the Pearson Correlation Coefficient as the final measurement method for this data set. There were two problems with this method. First, the resulting correlation coefficient does not offer a direct phenotypic description. Pearson Correlation Coefficients ranged from -1 to 1, with 1 being perfect correlation. The absolute differences between scores for phenotypic extremes was quite small in magnitude: assembly-prone domains had a Pearson Correlation Coefficient of around 0.8, while the non-assembling domain would have a coefficient of around 0.4. A coefficient of 0.4 suggests significant correlation, despite no detectable enrichment in stress granules. The likely basis for this positive correlation is that both the stress granule marker and the cPrLDs/sPrLDs are excluded from the vacuole. This highlights the second limitation of the method: we were interested in the fraction of cytoplasmic PrLD that localizes to stress granules, but the cytoplasm only occupies a small fraction of the cell. Exclusion of both fluorophores from the vacuole artificially inflates the correlation coefficient, but since the size of the vacuole varies substantially, some of the variation in correlation coefficients could be attributable to differences in vacuole size. Additionally, different PrLDs show different degrees of nuclear enrichment; enrichment of a PrLD in the nucleus would artificially lower the correlation coefficient. Collectively, these results suggested the need for a method that would specifically quantify the cytoplasm, and that would report a value that was more directly interpretable.

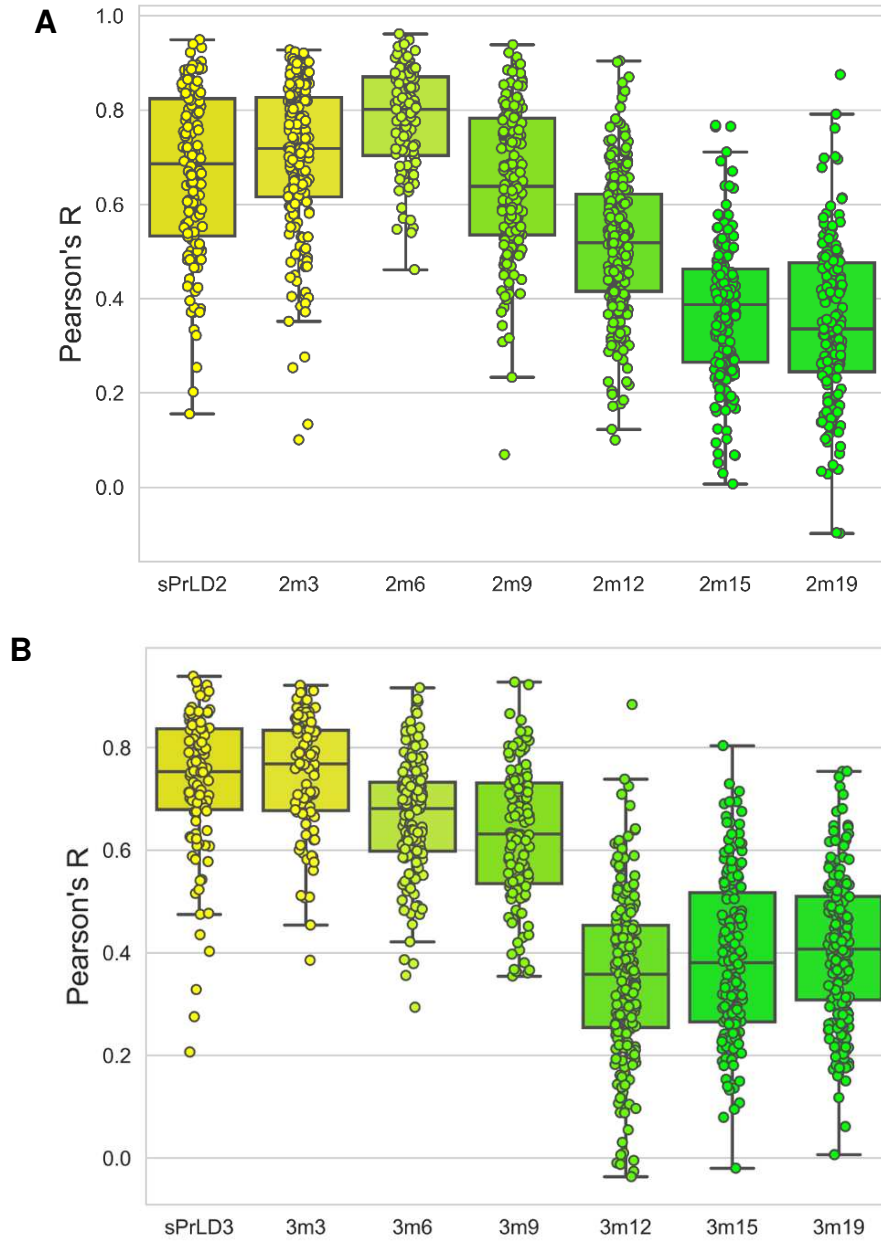


Figure 3.1: (A) Quantification of sPrLD2 & sPrLD2 aliphatic deletions after heat shock. 2m3 refers to sPrLD2 minus 3 I, L, or V residues, 2m6 is sPrLD2 minus 6 I, L, or V residues, etc. Each dot is the correlation coefficient for a single cell in an image. Median Pearson Correlation Coefficients tend to decrease as the domains become less aliphatic, or correlation with a stress granule marker (Pab1) decreases as assembly propensity decreases. **(B)** Quantification of sPrLD3 & sPrLD3 aliphatic deletions following the same naming convention on the X axis as in (A). Correlations are between GFP-sPrLDS & Pab1-mCherry. Image analysis carried out as described in Materials & Methods.

Partition Coefficient

One of the challenges with identifying the cytoplasm was the lack of clear pixel intensity differences between the vacuole, and the cytoplasm when labeling sPrLDs and Pab1. Unless we had diffuse sPrLD in the cytoplasm, the exact boundary between the vacuole and the cytoplasm of the cell was unclear to the eye, and even more elusive to describe quantitatively. Pixel intensities in the “dark” cytoplasm had no significant difference in number value from the pixel intensities in the “dark” vacuole. With the inability to clearly distinguish cytoplasm pixels from vacuole pixels in cells, the amount of cytoplasm in each cell can be either under- or overestimated, leading to a level of “noise” in the data we sought to remove.

In order to define the partitioning of sPrLDs between the bulk cytoplasm and stress granules, we first needed to be able to define the cytoplasm. To identify cytoplasm pixels, we decided to illuminate the cytoplasm by tagging Rpl1b, a yeast 60S ribosomal subunit protein [36]. We endogenously labeled Rpl1b with mTagBFP2, a monomeric blue fluorescent protein [38]. Endogenous labeling was carried out as described in Materials & Methods.

Importantly, exclusion of fluorophores from the nucleus and vacuole meant clear intensity differences between the cytoplasm and these organelles. Thus, we were able to demonstrate that the cytoplasm and stress granules in the cell were clearly labeled with fluorophores, and that minimal bleed through into the nucleus and vacuole was detected (Figure 3.2). With each region in the cell now illuminated, the cell pixels could be identified quantitatively because they exceed a fluorescence intensity threshold.

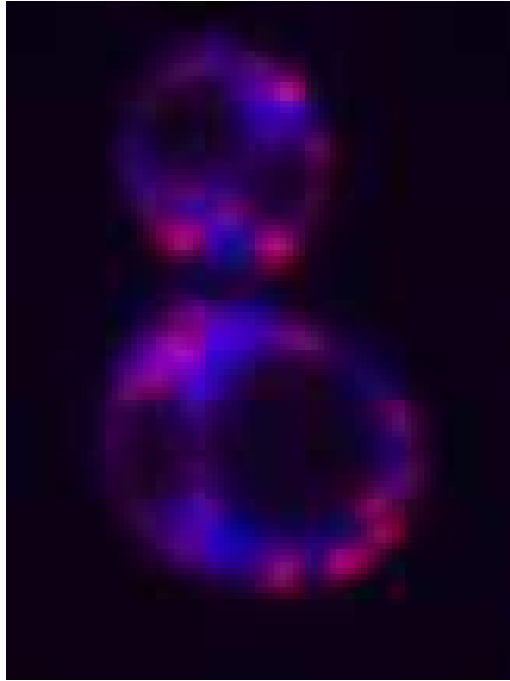


Figure 3.2: Yeast cells after heat shock expressing Rpl1b-mTagBFP2, and Pab1-mCherry to label the cytoplasm and stress granules, respectively. Each cell is displaying a nucleus and a vacuole devoid of significant fluorescence intensity relative to the labeled regions. Accurate isolation of blue “cytoplasm” pixels is now achievable thorough selection of pixels exceeding a fluorescence intensity threshold in the blue channel. Cell pixels are parsed into either cytoplasmic pixels, or stress granule pixels based on exceeding a threshold fluorescence intensity, and the GFP intensity distribution between the two pixel types can then be determined.

Once we were able to define cytoplasmic pixels and stress granule pixels, we were able to introduce a different measurement coefficient for quantification. We proposed creating a ratio: dividing the median GFP pixel intensity in stress granule pixels by the median GFP pixel intensity in cytoplasmic pixels (Figure 3.3 A). This would allow us to accurately describe the distributions of GFP intensities that arose from different assembly propensity phenotypes. In conditions of diffuse GFP-sPrLD after heat shock, the median GFP intensity in both compartments (cytoplasm and stress granules) is roughly equal, and would have a Partition Coefficient close to a value of one. When assembly propensity is high, like with the full-length sPrLDs, GFP signal is more concentrated in stress granules after heat shock. This results in a Partition Coefficient exceeding a value of one (Figure 3.3 B).

A

$$\frac{\text{Median GFP intensity in stress granules}}{\text{Median GFP intensity in cytoplasm}} = \text{Partition Coefficient}$$

B

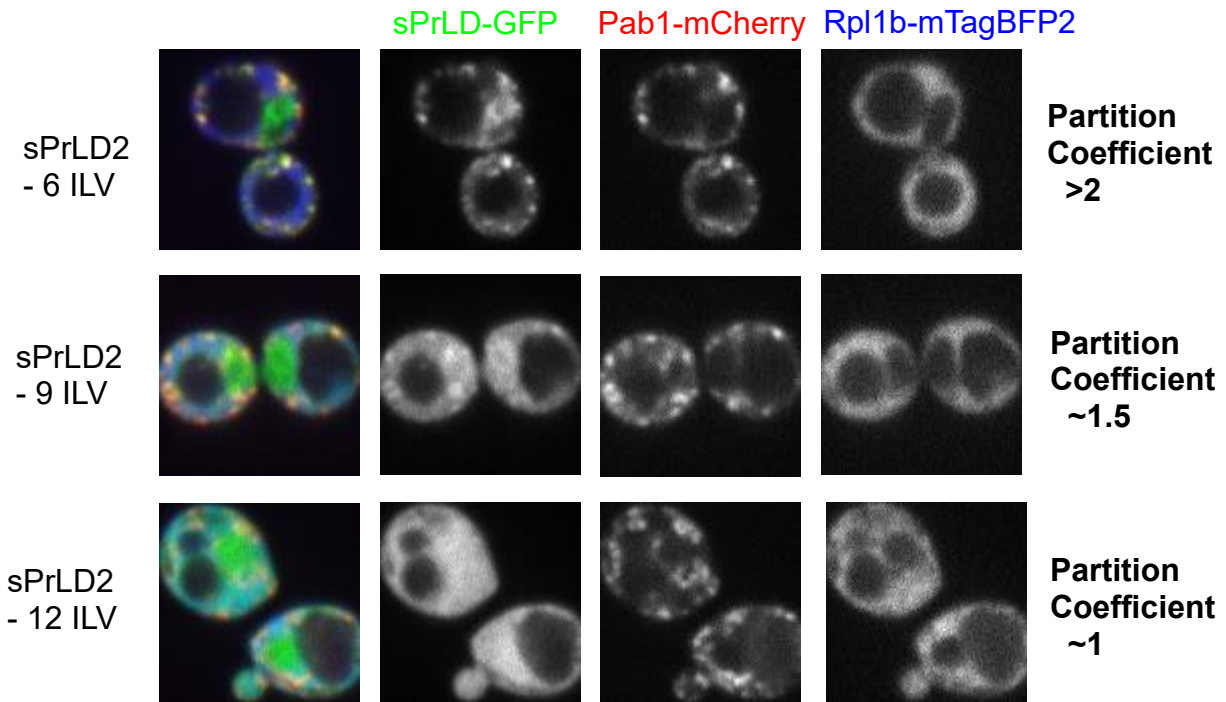


Figure 3.3: (A) Equation for the partition coefficient measurement to describe colocalization with a stress granule marker. The median GFP intensity in each compartment of the cell is divided to reveal the degree of GFP enrichment in stress granules. **(B)** Cytoplasmic labeling (Rpl1b) shown with sPrLD and stress granule (Pab1) labeling for three different sPrLD assembly propensities. GFP granules (puncta) in sPrLD2 -6 ILV indicates high assembly propensity, diffuse GFP in sPrLD2 -12 ILV indicates low assembly propensity. Partitioning into stress granules relative to cytoplasm reliably reflects assembly propensity of sPrLDs after heat shock.

Finally, to account for background fluorescence, we took images of an area not containing any cells (out-of-plane images) and subtracted that “blank” or “background”

image from each image we quantify. This background image reflects the autofluorescence and scattering of light by buffer and media components. Subtraction of this image from images containing cells allows for more accurate reflection of the GFP enrichment in lower expressing cells. This background subtraction step also helped to correct for differential illumination of specimens in the field-of-view, a problem we had noticed on our microscope.

With this new means of analysis, we set out to quantify the entire assembly propensity gradient we had created in chapter 2. The results from quantification using the Partition Coefficient revealed standard curves of assembly propensity. Phase separation was largely characterized in a binary manner before introducing this type of analysis. Here, we had demonstrated measurable yet subtle changes in assembly propensity. Inherent to these demonstrations as well was measurement of Partition Coefficients for the transition range constructs, or constructs teetering on the edge of assembly. Measurement of these teetering domains is significant because they are sensitive to subtle changes in assembly behavior in cells. Using our Partition Coefficient paired with sPrLDs, we could quantitatively measure subtle changes in assembly across several teetering PrLDs in live cells (Figure 3.4). With these tools, we could now probe for changes in assembly behavior of the teetering sPrLDs.

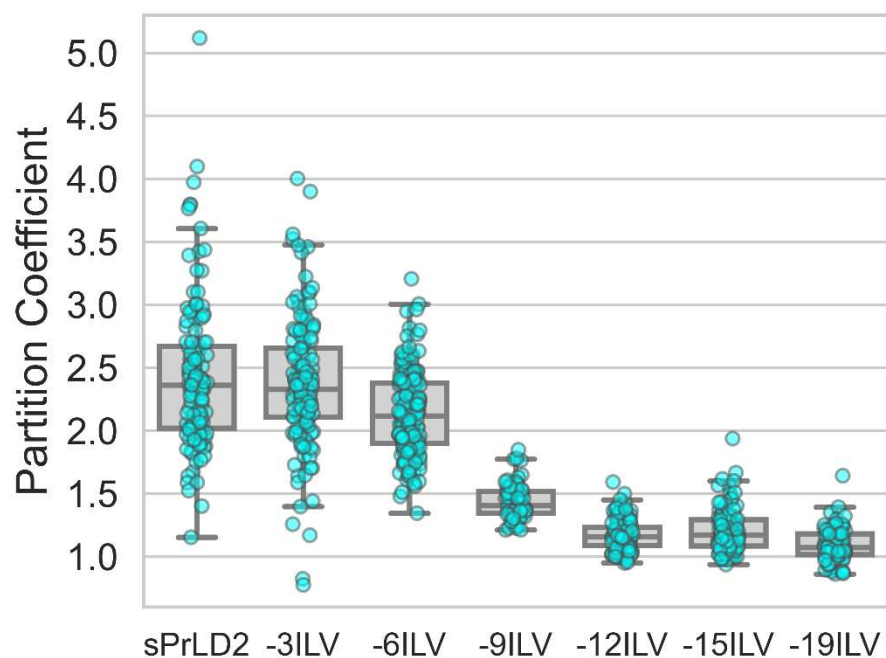
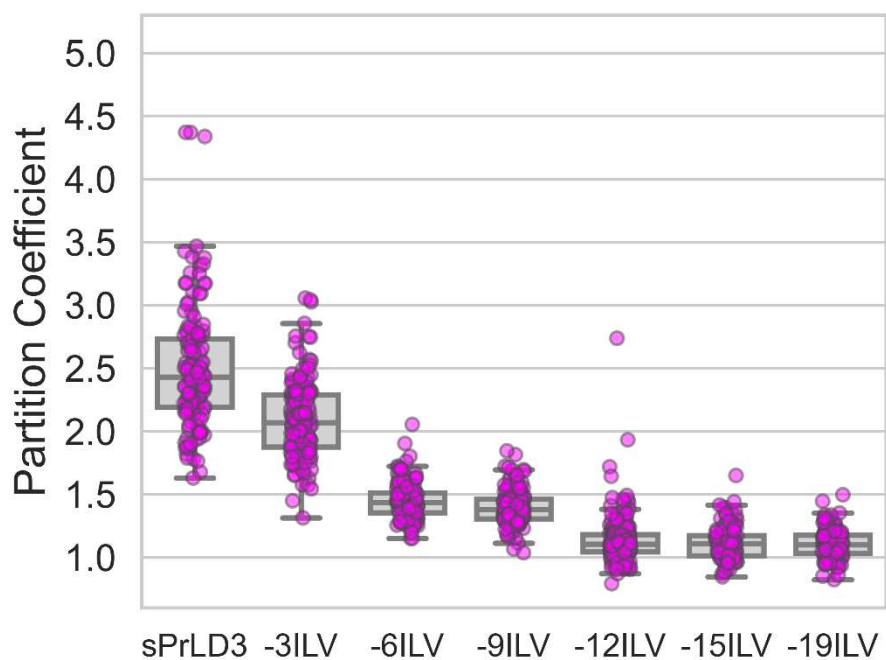
A**B**

Figure 3.4: (A) Standard curve of assembly propensity for sPrLD2 & sPrLD2 deletions. Each dot represents a single cell calculation of Partition Coefficient after heat shock. **(B)** Standard curve of assembly propensity for sPrLD3 & sPrLD3 deletions. Transition range constructs contain the inflection point for each set of sPrLDs in the assembly propensity gradient, and are sensitive to subtle changes in assembly.

Discussion

Once we could accurately and reproducibly measure partitioning, or the degree of colocalization of sPrLDs with a stress granule marker, we began to focus on a fundamental question regarding the mechanism of PrLD recruitment to stress granules. When we revealed the inflection points in the assembly propensity curves, we thought we could use those inflection points to measure changes in the assembly behavior of sPrLDs. Thus, we set out to test the concentration dependence for recruitment of PrLDs to stress granules by expressing them under stronger promoters.

In chapter 1 we saw that PrLDs are classically associated with concentration-dependent homotypic interactions when recruited to solid-state assemblies. We also saw in chapter 1 that PrLDs can be recruited to liquid-like assemblies through heterotypic interactions. An unanswered question in the field we hoped to address was whether the heterotypic interactions that recruit PrLDs to stress granules displayed a simple concentration-dependence as well. To address this question in chapter 4, we increased expression of sPrLDs to higher concentrations to see if we would observe an increase in Partition Coefficients. Our hypothesis was that increases in Partition Coefficients upon expression under a stronger promoter would indicate a simple concentration dependence. Thus, measuring the effect that increased concentration of PrLDs has on PrLD recruitment to stress granules will reveal if there is a simple concentration-dependence for heterotypic interactions.

Materials and Methods

Endogenous labelling of mTagBFP2

The *Saccharomyces cerevisiae* strain YER1405 was diluted to an Optical Density=0.1 in 100mL YPAD, and grown to an OD between 0.4 and 0.8. Cells were then permeabilized with TE-LiAc before being transformed through cloning by recombination. mTagBFP2 was amplified by PCR to contain tails with homology to RPL1B. mTagBFP2 was inserted in-frame just before the stop codon of RPL1B with the KanMX selectable marker. Cells were selected through plating on YPAD+G418 to select for KanMX positive mutants.

Strains & Growth Conditions

The experiments discussed in this chapter were carried out using the model organism *Saccharomyces cerevisiae* grown in standard yeast media, and with standard growth conditions [32]. Plasmids were transformed into the yeast strain YER2453 (MATa his3D1 leu2D0 met15D0 ura3D0 PAB1-mCherry::URA3 RPL1B::RPL1B-mTagBFP2::KanMX) using the standard techniques detailed in [33]. YER2453 is a derivative of YER1405. Yeast were grown at 30°C in pre-stress conditions, or heat shocked at 46°C for 30 minutes before imaging.

Stress Conditions

Cells were grown to midlog phase at 30 °C in SC-Leu media (to select for pER843) prior to stress induction. For heat shock, 1 mL of cells was concentrated to 50 µL and incubated in a 46 °C water bath for 30 min prior to imaging.

Confocal Fluorescence Microscopy

Imaging was performed on an Olympus (IX83) Inverted Spinning Disk Confocal Microscope using a 100× objective, and a 488nm laser along with a 561nm laser. All images were captured as single planes.

Pearson Correlation Coefficient analysis

Cells were imaged by DIC C, a 488nm laser, and a 561nm laser. Image segmentation and quantification was carried out as described in appendix (see *Pearson Correlation Coefficient analysis*)

Partition Coefficient analysis

Cells were imaged by DIC C, a 405nm laser, 488nm laser, and a 561nm laser. Image segmentation and quantification was carried out as described in appendix (see *Partition Coefficient analysis*)

CHAPTER 4: PROBING THE CONCENTRATION DEPENDENCE FOR PRION-LIKE DOMAIN RECRUITMENT TO STRESS GRANULES

Introduction

In chapter 3 we developed a method for quantifying the Partition Coefficient of PrLDs between the bulk cytoplasm and stress granules. Using the Partition Coefficient, we generated standard curves relating the assembly propensity to aliphatic content for sPrLDs. In this chapter, we asked whether this standard curve is shifted by changes in levels of PrLD expression, as would be expected if recruitment to stress granules was concentration dependent.

The standard curves generated in chapter 3 were quantified using the yeast *SUP35* promoter for expression of sPrLDs. This promoter provided a moderate level of expression as a baseline. It is naturally optimized to keep expression levels from creeping up too high due to aggregation of Sup35 resulting from over-expression [38]. To probe the concentration-dependence for recruitment of PrLDs, we compared recruitment of sPrLDs expressed from the *SUP35* promoter to sPrLDs expressed from two strong constitutive promoters: *GPD* and *ADH1*.

Results

We first tested a number of constitutive promoters. To do this, we expressed sPrLD2 – 6 ILV, and sPrLD2 -9 ILV under 6 promoters in addition to *SUP35* to verify that the level of expression of sPrLDs was consistent with published results, and results obtained by other members of the lab using the same promoter set. The promoter set expression levels were analyzed by western blot to verify that expression levels of these transition range constructs under control of the different promoters was higher than with

the *SUP35* promoter. Importantly, we saw expression levels for these stronger promoters exceeding the detectable levels of expression of sPrLDs under control of the *SUP35* promoter. (Figure 4.1) Two promoters from this set were chosen for expression of sPrLDs at higher levels than with the *SUP35* promoter. Very high expression levels were observed with the *GPD* promoter, and moderately high expression was observed for the *ADH1* promoter (Figure 4.1).

Further support for selecting these promoters came from a 2018 publication where Xiong et al. characterized the expression levels of multiple constitutive promoters in yeast in conditions of heat stress [40]. In stress induced by heat shock at 39°C, expression of GFP at 8 hours for the *TDH3* (*GPD*) and *ADH1* promoters was detected at around 6, and 2 thousand arbitrary units, respectively above background fluorescence. The *TDH3* (*GPD*) promoter, the highest expressing constitutive promoter in the yeast genome, showed the highest detectable levels of GFP under heat stress conditions at 8 hours, with consistently high but slightly attenuated expression of GFP for *ADH1* relative to *GPD* [39].

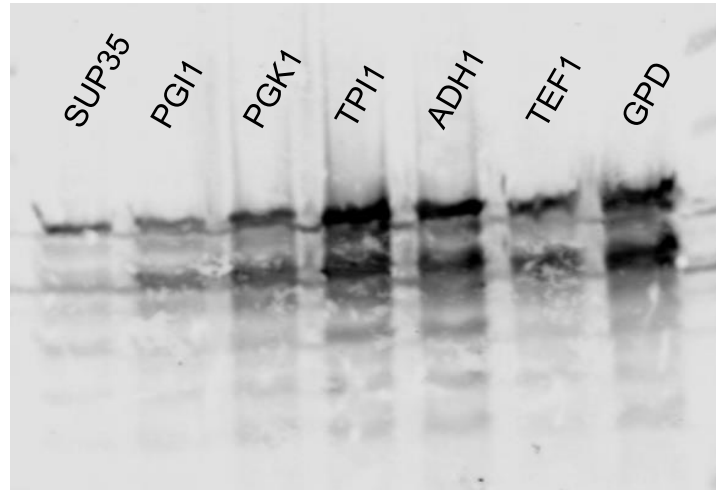
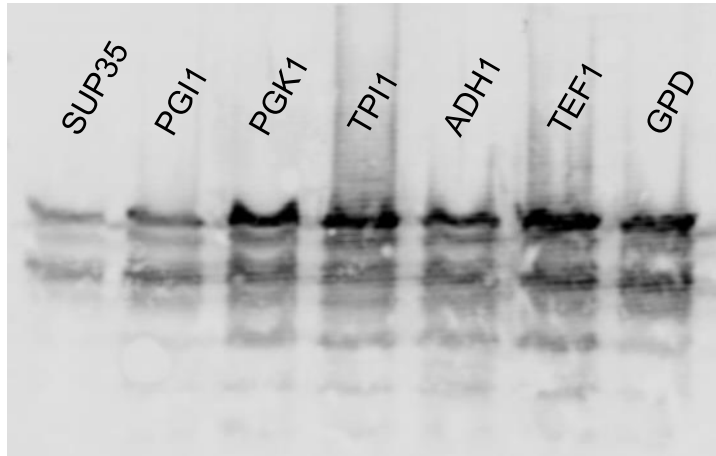
A**B**

Figure 4.1: **(A)** Western blot of sPrLD2 -6 ILV under control of seven different promoters. Cells were grown to equal density as measured by absorbance at OD₆₀₀ before harvest, lysis, and detection with anti-GFP. **(B)** Western blot of sPrLD2 -9 ILV under control of seven different promoters. Cells were subject to the same blot conditions as in (A).

Once we had verification that the expression levels of the transition range constructs was indeed higher than the levels obtained with the SUP35 promoter, we tested the highest expressing constructs with our Pearson Correlation Coefficient analysis method. To do this, we imaged all of the sPrLDs in the assembly propensity gradient under control of the SUP35 promoter, as well as the transition range constructs under control of the GPD promoter. If there were a simple concentration-dependence, we would be most likely to detect this with the highest levels of expression first, so the GPD promoter was a natural first choice. With the SUP35 promoter constructs, we imaged each domain in each curve on three separate days, in a random order of progression through the dataset. In addition to the SUP35 promoter constructs along the entire curve, we also tested the transition range constructs under control of the GPD promoter on three separate days per domain across both curves. We subsequently analyzed the images using the Pearson Correlation Coefficient method of analysis described in chapter 3. From this analysis, we found preliminary evidence for a modest concentration dependence for PrLD recruitment to stress granules. (Figure 4.2)

This preliminary evidence was encouraging, so we transformed the sPrLDs into our cytoplasm-labeled strain described in chapter 3 (YER2453) to perform Partition Coefficient analysis. To verify the modest concentration-dependence for recruitment of PrLDs we had observed using the Pearson Correlation Coefficient method would require a more rigorous approach. To examine the heterotypic interactions more deeply, we employed use the ADH1 promoter as a “middle” level of expression between SUP35 and GPD to see if we could detect any changes in recruitment of the teetering constructs at these three different levels of expression.

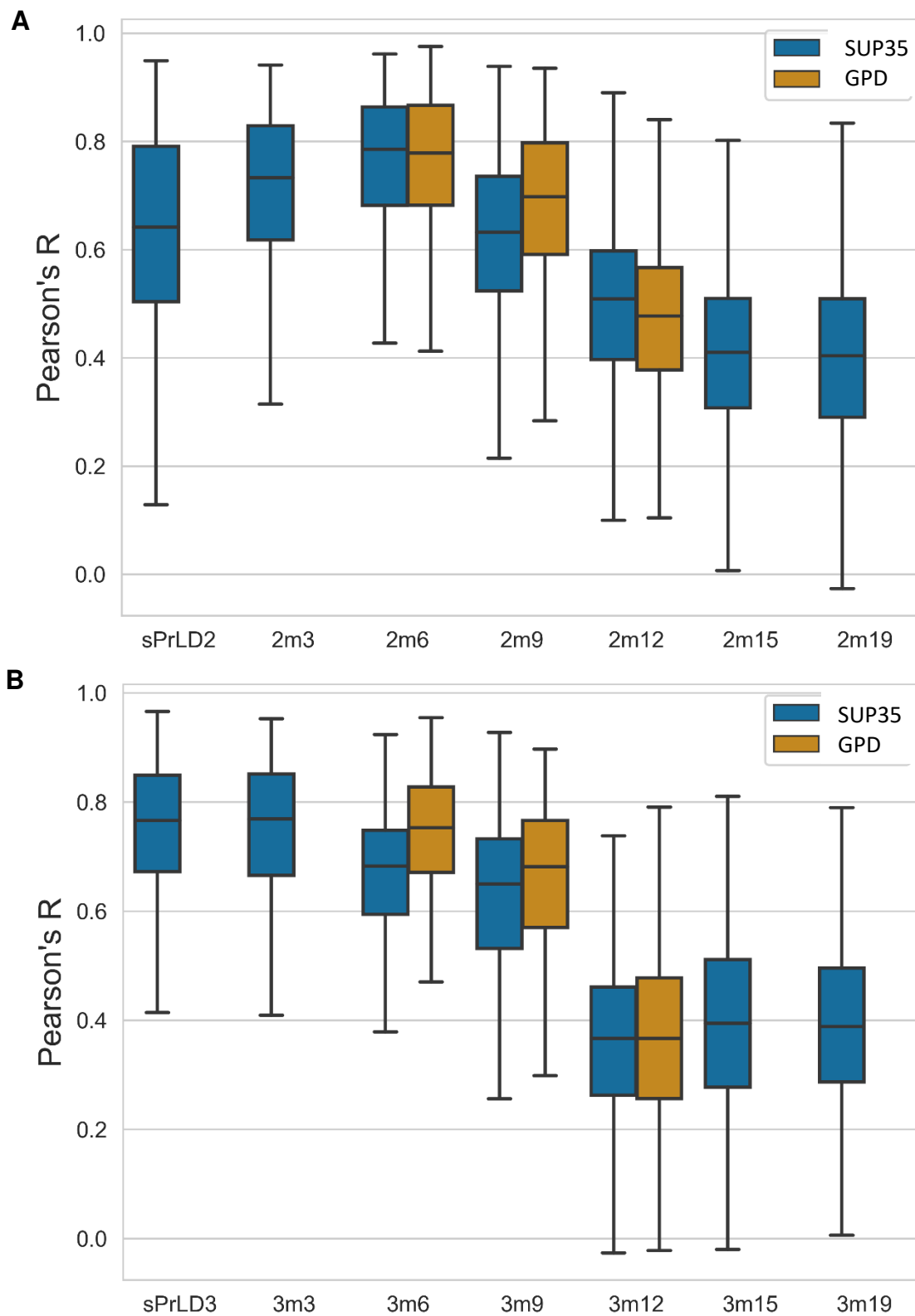


Figure 4.2: (A) Pearson Correlation Coefficient statistics for sPrLD2 & sPrLD2 aliphatic deletions under control of the SUP35 promoter displayed along with statistics for the

transition range constructs under control of the GPD promoter. Statistics are from 3 replicate populations of cells from three different days of imaging per construct, per promoter. A slightly increased median Pearson Correlation Coefficient for sPrLD2 -9ILV (sPrLD2m9) suggests a modest concentration dependence for assembly of this domain. **(B)** Pearson Correlation Coefficient statistics for sPrLD3 & sPrLD3 aliphatic deletions under control of the SUP35 promoter displayed along with statistics for the transition range constructs under control of the GPD promoter. Statistics are from 3 replicate populations of cells from three different days of imaging per construct, per promoter. A slightly increased median Pearson Correlation for the sPrLD3 -6 ILV & sPrLD3 – 9ILV further supports a modest concentration dependence for assembly of PrLDs to stress granules.

The *ADH1* promoter was added in the next round of testing because a modest concentration-dependence for recruitment may present with a modest increase in Partition Coefficient. However, if it scored with a higher Partition Coefficient than with *GPD*-promoter-based expression, the results would be difficult to interpret. Testing of the constructs under each promoter was carried out as described in chapter 3. We measured Partition Coefficients first for *SUP35* promoter-based expression of sPrLDs as a set of baseline points. We then analyzed images acquired from *GPD*- and *ADH1*-based-expression to test for the presence of a modest concentration dependence in the teetering constructs.

We planned to compare the median Partition Coefficients between constructs and promoter type for a given round of testing in the final analysis. We first needed to ensure that we could obtain reproducible median Partition Coefficients when imaging with a given promoter. To address this, the *SUP35*-based expression standard curve data collection and quantification was carried out in triplicate to ensure consistent median Partition Coefficients at this lowest expression level (Figure 4.3). Once we had data demonstrating reproducibility of the Partition Coefficient analysis technique on the *SUP35* promoter data, we moved forward with data collection using the *GPD* and *ADH1* promoter constructs.

Quantification of images obtained from expression of sPrLDs under the stronger promoters revealed that the median Partition Coefficients for some of the transition range constructs showed a modest increase when expressed under the *GPD* promoter. The median coefficients tended to be lowest for the *SUP35* constructs, and *ADH1* tended to have median Partition Coefficients directly between the strongest and weakest promoter scores (Figure 4.4, 4.5). This scoring hierarchy is apparent in Figure 4.5, where sPrLD3 -6 ILV & sPrLD3 -9 ILV display the ascending order of Partition Coefficients based on expression level. Importantly, the increase in median coefficients is less than the increase in median coefficients when slightly modifying the aliphatic residue content of the domain. This supported the hypothesis that there is a modest concentration dependence for recruitment of PrLDs to stress granules. However, when compared to the magnitudes of change in Partition Coefficient observed when deleting aliphatic residues in sPrLDs, higher concentrations of PrLD appear to have a much less significant impact on the recruitment of PrLDs to stress granules (Figure 4.4, 4.5).

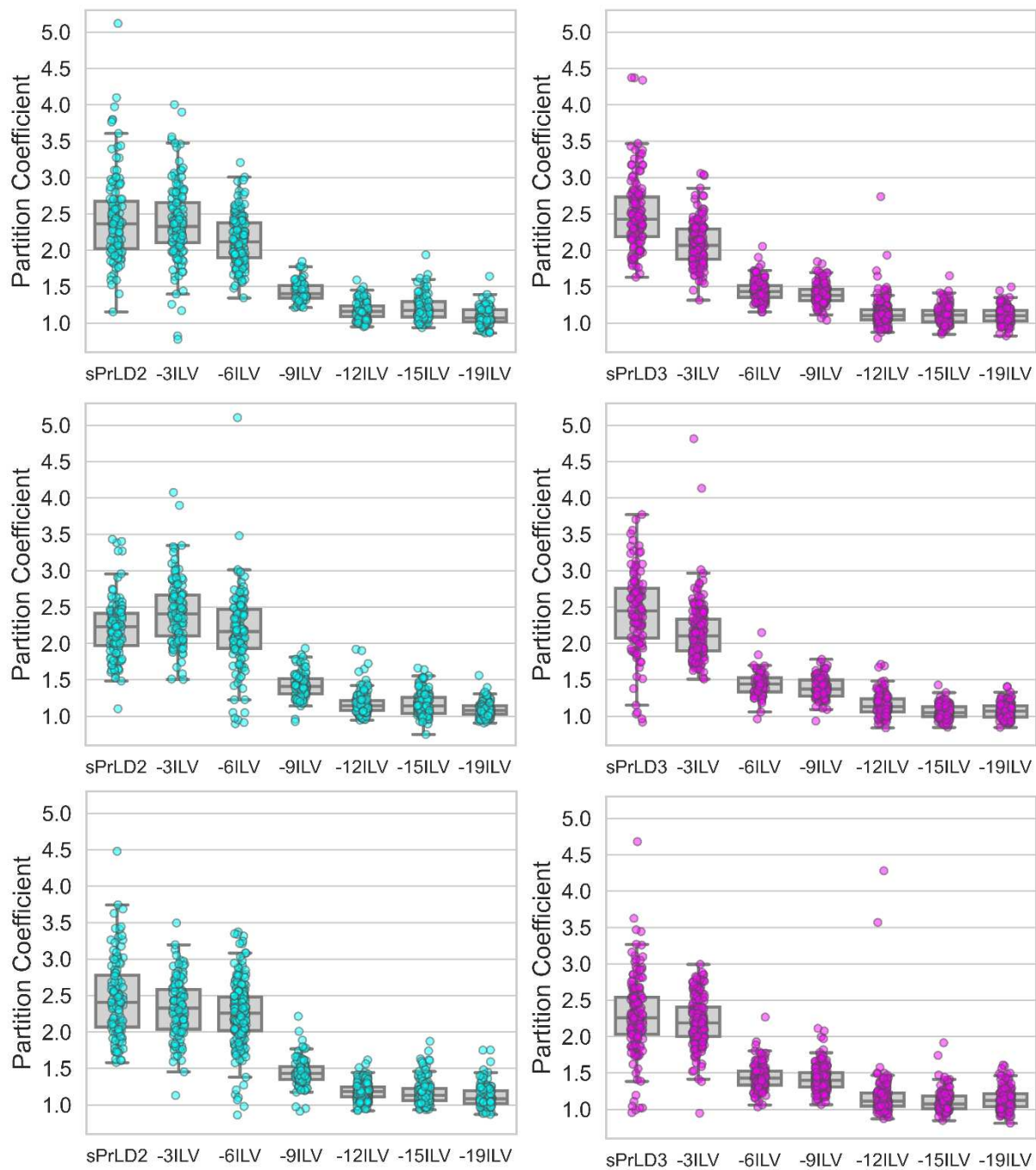


Figure 4.3: Three replicate measurements of Partition Coefficients for expression of two sPrLDs under the *SUP35* promoter. Three rounds of imaging were performed for sPrLD2 & sPrLD2 deletions in a random order each time (cyan dots). In addition to this, three rounds of imaging were performed for sPrLD3 & sPrLD3 deletions in a random order each time (magenta dots).

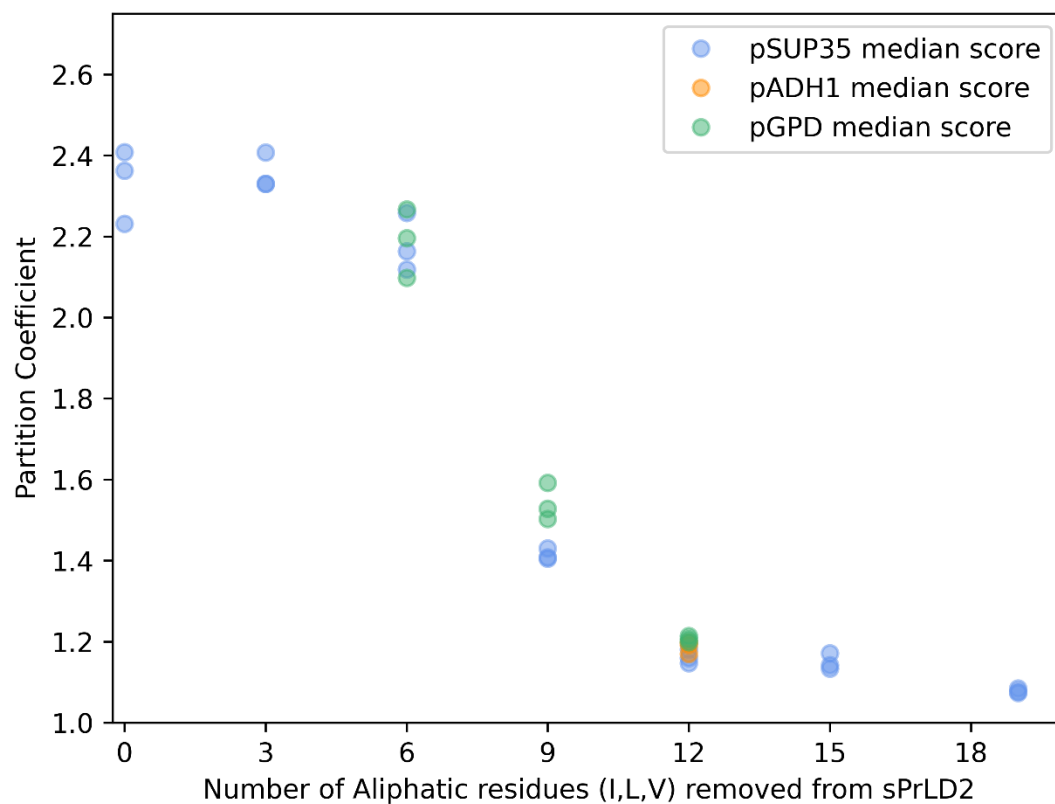


Figure 4.4: Median Partition Coefficients for sPrLDs at three expression levels. This data is incomplete at this time. Triplicate rounds of testing were carried out for p*SUP35* for every sPrLD in the set. p*GPD* was tested in only one of the transition range constructs so far. Testing of the rest of the transition range constructs with p*ADH1* and p*GPD* will follow, along with testing of “-3 ILV” & “-15 ILV” at each increased expression level to complete this figure.

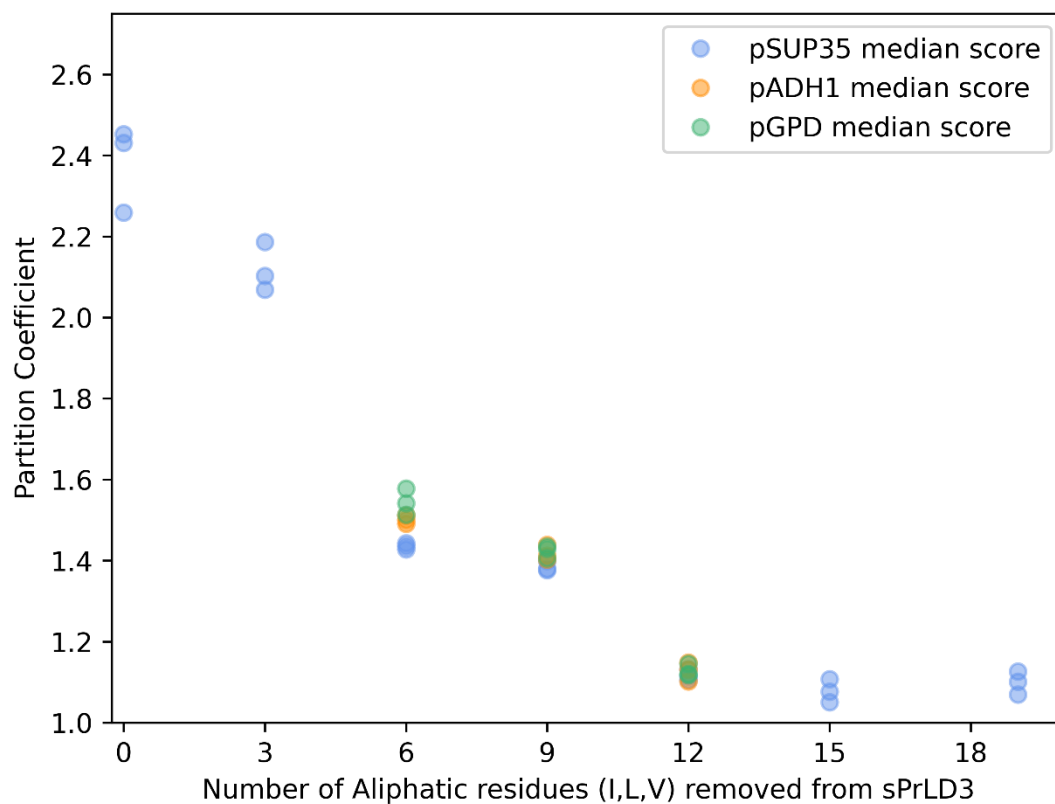


Figure 4.5: Median Partition Coefficients for sPrLDs at three expression levels. This data is incomplete at this time. Triplicate rounds of testing were carried out for p*SUP35*, p*ADH1*, and p*GPD* in the transition range constructs so far. Testing of the “-3 ILV” & “-15 ILV” at each increased expression level will be carried out to complete this figure.

Discussion

Concentration appears to play an important role in phase separation in cells because of data collected *in vitro* [41], [42]. However, protein concentrations for many *in vitro* phase separation studies often exceed one or two orders of magnitude above the relevant physiological concentrations [41]. Until now, there had only been early indications from work done *in vivo* that concentration plays a much more modest effect on the assembly of some types of LLPS-prone domains, such as those recruited to the nucleolus [31]. Now, with quantification of the data from expression of sPrLDs under different strength promoters, PrLDs that are recruited to stress granules also display a modest concentration dependence for recruitment. Using the Partition Coefficient method of analysis we discussed in chapter 3, we showed that small (<15%) increases in median Partition Coefficient result from expression under the *GPD* promoter, the strongest constitutive promoter in the yeast genome. To us, this supports the hypothesis that a modest concentration dependence is observed for PrLDs recruited to stress granules as well. Under control of two different strength promoters (*SUP35* vs *GPD*), changes in partition coefficient were less significant than changes in aliphatic content of the domains.

This supports the hypothesis that composition drives recruitment to stress granules primarily through heterotypic interactions. Increasing PrLD concentration by using a stronger promoter should increase the partitioning of PrLDs into stress granules if there is a simple concentration-dependence. However, we did not observe a

substantial increase in partitioning that would support the concentration-dependence for recruitment of PrLDs to stress granules.

With an excess of interaction partners for heterotypic interactions between PrLDs and components in stress granules, recruitment appears to proceed at similar rates from a variety of concentrations of PrLD. Partitioning is only modestly increased when PrLD concentration is increased by expression with the strongest naturally occurring constitutive promoter in yeast. This may also suggest that homotypic interactions only play a small role in the recruitment of PrLDs to stress granules in yeast. This is critically important as concentration-dependent, homotypic interactions are classically associated with yeast PrLDs. Yet, PrLDs that are recruited to stress granules continue to display divergent features from classical PrLDs. These unique features include a necessary and sufficient aliphatic residue content for assembly seen in chapter 2, and a modest concentration dependence for recruitment to stress granules seen in this chapter. We believe the unique features of PrLDs recruited to stress granules characterized in this work are early indications of a need for defining a separate class of Prion-like Domains. PrLDs that are recruited to stress granules differ in their compositional preferences, and their mechanism of recruitment. Additionally, they also display a modest concentration dependence for assembly, as opposed to the simple concentration dependence observed for classical PrLDs.

Table 4.1 Partition Coefficient median values and sample sizes for 3 promoters

SUP35 PROMOTER						
Construct	Round One Medians	sample size (n1)	Round Two Medians	sample size (n2)	Round Three Medians	sample size (n3)
sPrLD2	2.4082	120	2.2314	127	2.3621	126
sPrLD2m3	2.3301	114	2.4070	126	2.3299	140
sPrLD2m6	2.2579	216	2.1637	131	2.1184	198
sPrLD2m9	1.4302	104	1.4083	106	1.4046	71
sPrLD2m12	1.1810	242	1.1471	154	1.1594	116
sPrLD2m15	1.1338	158	1.1420	141	1.1715	105
sPrLD2m19	1.0845	137	1.0768	130	1.0736	95
sPrLD3	2.2589	168	2.4523	114	2.4311	143
sPrLD3m3	2.1866	183	2.1023	189	2.0689	156
sPrLD3m6	1.4281	135	1.4427	63	1.4356	129
sPrLD3m9	1.3991	264	1.3762	117	1.3794	146
sPrLD3m12	1.1147	154	1.1318	113	1.1062	211
sPrLD3m15	1.0764	122	1.0508	136	1.1073	119
sPrLD3m19	1.1265	173	1.0699	118	1.1012	126

ADH1 PROMOTER						
Construct	Round One Medians	sample size (n1)	Round Two Medians	sample size (n2)	Round Three Medians	sample size (n3)
sPrLD2						
sPrLD2m3	-	-	-	-	-	-
sPrLD2m6						
sPrLD2m9						
sPrLD2m12	1.1985	170	1.1916	150	1.1696	131
sPrLD2m15	-	-	-	-	-	-
sPrLD2m19						
sPrLD3						
sPrLD3m3	-	-	-	-	-	-
sPrLD3m6	1.5003	126	1.4909	115	1.5112	112
sPrLD3m9	1.4027	106	1.4124	128	1.4389	220
sPrLD3m12	1.1018	106	1.1487	83	1.1314	148
sPrLD3m15	-	-	-	-	-	-
sPrLD3m19						

GPD PROMOTER						
Construct	Round One Medians	sample size (n1)	Round Two Medians	sample size (n2)	Round Three Medians	sample size (n3)
sPrLD2						
sPrLD2m3	-	-	-	-	-	-
sPrLD2m6	2.2670	145	2.1958	112	2.0975	98
sPrLD2m9	1.5027	81	1.5284	123	1.5924	72
sPrLD2m12	1.2144	128	1.1992	135	1.2059	80
sPrLD2m15	-	-	-	-	-	-
sPrLD2m19						
sPrLD3						
sPrLD3m3	-	-	-	-	-	-
sPrLD3m6	1.5417	73	1.5775	148	1.5132	106
sPrLD3m9	1.4353	106	1.4295	132	1.4059	109
sPrLD3m12	1.1176	93	1.1201	99	1.1450	105
sPrLD3m15	-	-	-	-	-	-
sPrLD3m19						

Materials and Methods³

Strains & Growth Conditions

The experiments discussed in this chapter were carried out using the model organism *Saccharomyces cerevisiae* grown in standard yeast media, and with standard growth conditions.[33] Plasmids were transformed into the yeast strain YER1405 (MATa his3D1 leu2D0 met15D0 ura3D0 PAB1-mCherry::URA3) using the standard techniques detailed in [33]. YER1405 is a derivative of BY4741 [32]. Yeast were grown at 30°C in pre-stress conditions, or heat shocked at 46°C for 30 minutes before imaging.

Plasmid construction

Generation of plasmid backbone for sPrLD expression under the SUP35 promoter (pER843) was carried out by inserting GFP between the HindIII and BamHI restriction sites following the *SUP35* promoter in the plasmid pJ526 from [33]. sPrLDs were then cloned into pER843 in frame with GFP at its N-terminus between the BamHI and BglII sites. Generation of the *GPD* and *ADH1* promoter plasmids was carried out by Kacy Paul, and was not included in the publication listed below.

Stress Conditions

Cells were grown to midlog phase at 30 °C in SC-Leu media (to select for the plasmids) prior to stress induction. For heat shock, 1 mL of cells was concentrated to 50 µL and incubated in a 46 °C water bath for 30 min prior to imaging.

Confocal Fluorescence Microscopy

Imaging was performed on an Olympus (IX83) Inverted Spinning Disk Confocal Microscope using a 100× objective, and a 405nm laser, a 488nm laser along, and a 561nm laser. All images were captured as single planes.

³This section is reformatted from the following publication: Boncella A.E., Shattuck J.E., Cascarina S.M., Paul K.R., Baer M.H., Fomicheva A, Lamb A.K, Ross E.D. (2020). Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proceedings of the National Academy of Sciences* 117 (11) 5826-5835. My contribution to that work was building and testing of the following constructs after heat shock: Δ FWY, Δ FWY \rightarrow ILV, Δ FWY \rightarrow QN. As well as imaging sPrLD2, sPrLD3, cPrLD1 after heat shock. Aliphatic deletion domains were cloned into pER843 using the same method listed in this section.

Western Blotting

Cells were grown to midlog phase at 30 °C in SC-Leu media (to select for the plasmids) before harvesting. Volumes collected for each culture were normalized to the lowest density culture, as assessed by OD600 measurements. Cells were harvested by centrifugation at 3,000 rpm for 5 min at 4 °C. Cell lysis was performed using a SUMEB lysis buffer. A volume of 15 µL of each sample was run on a polyacrylamide gel and then transferred onto a PVDF membrane. GFP-PrLD fusions were probed using an anti-GFP antibody (Santa Cruz Biotechnology). Images were acquired using Li-Cor Odyssey Clx imaging system.

Image Analysis

Images were exported from SlideBook® as 16-bit TIFFS. Image masks were generated using YeastSpotter® to identify cells in DIC C images that were exported along with fluorescence image stacks. Each three-color fluorescence image taken was exported as three separate single channel TIFF files to feed in to in-house built Python scripts to perform Partition Coefficient analysis. Background fluorescence subtraction was carried out to minimize effects brought on by low expression. Partition coefficient analysis was carried out by isolating pixels in images that were illuminated by fluorescently labeled cytoplasmic or stress granule components to identify both compartments in cells, followed by quantifying the GFP intensity distribution in each compartment of the cell.

Pearson Correlation Coefficient analysis

Cells were imaged by DIC C, a 488nm laser, and a 561nm laser. Image segmentation and quantification was carried out as described in appendix (see *Pearson Correlation Coefficient analysis*)

Partition Coefficient analysis

Cells were imaged by DIC C, a 405nm laser, 488nm laser, and a 561nm laser. Image segmentation and quantification was carried out as described in appendix (see *Partition Coefficient analysis*)

⁴Part of this section is reformatted from the following publication: Boncella A.E., Shattuck J.E., Cascarina S.M., Paul K.R., Baer M.H., Fomicheva A, Lamb A.K, Ross E.D. (2020). Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proceedings of the National Academy of Sciences* 117 (11) 5826-5835. My contribution to that work was building and testing of the following constructs after heat shock: Δ FWY, Δ FWY \rightarrow ILV, Δ FWY \rightarrow QN. As well as imaging sPrLD2, sPrLD3, cPrLD1 after heat shock. Aliphatic deletion domains were cloned into the plasmids using the same method listed in this section.

CHAPTER 5: CONCLUSIONS

From the characteristics of solid-state assemblies we discussed in chapter 1, we saw why amino acid composition drives homotypic interactions of PrLDs into yeast prions despite primary sequence. These domains participate in repeated, in register, binding events that add to the fibril in the same manner each time. The shared features of solid-state assemblies also helps to explain the lack of enrichment for compositional elements such as charged residues, due to like charges repelling one another. Solid-state assemblies share a similar concentration dependence for recruitment of PrLDs through homotypic interactions whether they are functional or pathogenic assemblies. Thus, the shared classification of prion-like domains due to recruitment to solid-state assemblies is appropriately applied based on the shared characteristics. However, this contrasts with PrLDs that are recruited to liquid-like assemblies, as they appear to be quite different in their shared features from classical PrLDs.

PrLDs recruited to liquid-like assemblies such as stress granules differ in compositional preferences, and they are recruited primarily through a different mechanism. PrLDs recruited to liquid-like assemblies participate in heterotypic interactions via liquid-liquid phase separation, and recruitment appears to proceed from a variety of concentrations without significantly changing PrLD partitioning. In chapter 4 we saw that expression of several PrLDs teetering on the edge of assembly with the strong GPD promoter displayed only a modest increase in partitioning of PrLDs into stress granules, supporting this modest concentration dependence hypothesis.

The dividing line between classical PrLDs and PrLDs recruited through LLPS (Hereafter, LLPS-PrLDs) was obscured in early work by highlighting a concentration-dependence for LLPS *in vitro* when characterizing LLPS-PrLDs. Simplified component systems failed to account for the reduction of interaction partners for PrLDs compared to the potential interactions normally found in cells. This increased the necessary contribution of homotypic interactions for recruitment of PrLDs. Subsequently, this illuminated a potentially false concentration-dependence for LLPS-PrLDs. This perceived feature of LLPS-PrLDs persisted until the publication by the Brangwynne lab highlighting the composition-dependent thermodynamics controlling recruitment of domains to the nucleolus [32]. This early evidence provided initial support for the presence of a modest concentration dependence for LLPS-PrLDs. It wasn't until we were able to demonstrate the insignificant increases to median Partition Coefficients observed when sPrLDs were expressed under stronger promoters that we were convinced this modest concentration dependence may be ubiquitous among LLPS-PrLDs. Increased expression through use of the GPD promoter across several sPrLDs that teeter on the edge of assembly led to a <15% increase in median Partition Coefficient at most, suggesting a modest concentration-dependence for recruitment of PrLDs to stress granules.

Prion-like domains recruited to liquid-like assemblies, or LLPS-PrLDs, have three key deviations from classical PrLDs. The first difference is that domains recruited to liquid-like assemblies are recruited through a different mechanism than those recruited to solid-state assemblies. Heterotypic interactions comprise the majority of interactions leading to recruitment of LLPS-PrLDs, although the presence of homotypic interactions

within the separated liquid phase cannot be entirely ruled out. We saw evidence for these homotypic interactions in increased expression of sPrLDs in chapter 4. The modest increase in partitioning was likely due to a modest increase in homotypic interactions from the increased concentration of LLPS-PrLDs. This difference in mechanism of recruitment confers distinct compositional characteristics to LLPS-PrLDs and is partially responsible for the second deviation from classical PrLDs. The preference for charged residues in domains recruited to liquid-like is the second unique characteristic among LLPS-PrLDs. Charged residues are largely prohibitive of formation of cross- β -structures, and the preference for charged residues in LLPS-PrLDs further supports the need for reclassification of this sub-type of PrLD. The third significant deviation LLPS-PrLDs display in contrast to classical PrLDs is the lack of a simple concentration-dependence for recruitment. In chapter 4, we demonstrated evidence against a simple concentration-dependence to support the need for reclassification of this sub-type of PrLD. Heterotypic interactions dictate that the interacting partners for LLPS-PrLDs in the cytoplasm are in vast excess. Recruitment can proceed from a variety of concentrations of LLPS-PrLD. Reclassification of these functionally distinct PrLDs will allow for better understanding of the currently underdiscussed differences between these two types of PrLDs. To understand the role of LLPS-PrLDs in diseases, it is important to clearly define their unique features from classical PrLDs. To progress toward a greater understanding of PrLDs recruited to liquid-like assemblies, we must classify them in a way that demonstrates their clear divergence from classical PrLDs. Thus, we propose reclassifying PrLDs that are recruited to liquid-like assemblies as LLPS-PrLDs to accurately portray the differences among PrLDs.

REFERENCES

- [1] C. A. Ross and M. A. Poirier, "Protein aggregation and neurodegenerative disease," *Nature Medicine*, vol. 10, no. S7, pp. S10–S17, Jul. 2004, doi: 10.1038/nm1066.
- [2] E. H. Koo, P. T. Lansbury, and J. W. Kelly, "Amyloid diseases: Abnormal protein aggregation in neurodegeneration," *Proceedings of the National Academy of Sciences*, vol. 96, no. 18, pp. 9989–9990, Aug. 1999, doi: 10.1073/pnas.96.18.9989.
- [3] R. Wickner, "[URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*," *Science*, vol. 264, no. 5158, pp. 566–569, Apr. 1994, doi: 10.1126/science.7909170.
- [4] B. S. Cox, "Ψ, A cytoplasmic suppressor of super-suppressor in yeast," *Heredity*, vol. 20, no. 4, pp. 505–521, Nov. 1965, doi: 10.1038/hdy.1965.65.
- [5] F. Lacroute, "Non-Mendelian Mutation Allowing Ureidosuccinic Acid Uptake in Yeast," *Journal of Bacteriology*, vol. 106, no. 2, pp. 519–522, 1971, doi: 10.1128/JB.106.2.519-522.1971.
- [6] D. C. Masison and R. B. Wickner, "Prion-Inducing Domain of Yeast Ure2p and Protease Resistance of Ure2p in Prion-Containing Cells," *Science*, vol. 270, no. 5233, pp. 93–95, Oct. 1995, doi: 10.1126/science.270.5233.93.
- [7] N. Sondheimer and S. Lindquist, "Rnq1: an epigenetic modifier of protein function in yeast," *Molecular Cell*, vol. 5, no. 1, pp. 163–172, Jan. 2000, doi: 10.1016/S1097-2765(00)80412-8.
- [8] E. D. Ross, U. Baxa, and R. B. Wickner, "Scrambled Prion Domains Form Prions and Amyloid," *Molecular and Cellular Biology*, vol. 24, no. 16, pp. 7206–7213, Aug. 2004, doi: 10.1128/MCB.24.16.7206-7213.2004.
- [9] E. D. Ross, H. K. Edskes, M. J. Terry, and R. B. Wickner, "Primary sequence independence for prion formation," *Proceedings of the National Academy of Sciences*, vol. 102, no. 36, Sep. 2005, doi: 10.1073/pnas.0506136102.
- [10] M. D. Michelitsch and J. S. Weissman, "A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions," *Proceedings of the National Academy of Sciences*, vol. 97, no. 22, pp. 11910–11915, Oct. 2000, doi: 10.1073/pnas.97.22.11910.
- [11] E. D. Ross, A. Minton, and R. B. Wickner, "Prion domains: sequences, structures and interactions," *Nature Cell Biology*, vol. 7, no. 11, pp. 1039–1044, Nov. 2005, doi: 10.1038/ncb1105-1039.

- [12] S. Alberti, R. Halfmann, O. King, A. Kapila, and S. Lindquist, "A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins," *Cell*, vol. 137, no. 1, pp. 146–158, Apr. 2009, doi: 10.1016/j.cell.2009.02.044.
- [13] K. Kapeli, F. J. Martinez, and G. W. Yeo, "Genetic mutations in RNA-binding proteins and their roles in ALS," *Human Genetics*, vol. 136, no. 9, Sep. 2017, doi: 10.1007/s00439-017-1830-7.
- [14] E. Boke *et al.*, "Amyloid-like Self-Assembly of a Cellular Compartment," *Cell*, vol. 166, no. 3, 2016, doi: 10.1016/j.cell.2016.06.051.
- [15] C. P. Brangwynne *et al.*, "Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation," *Science*, vol. 324, no. 5935, pp. 1729–1732, Jun. 2009, doi: 10.1126/science.1172046.
- [16] H. J. Kim *et al.*, "Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS," *Nature*, vol. 495, no. 7442, pp. 467–473, Mar. 2013, doi: 10.1038/nature11922.
- [17] R. B. Wickner, H. K. Edskes, M. Son, S. Wu, and M. Niznikiewicz, "How Do Yeast Cells Contend with Prions?," *International Journal of Molecular Sciences*, vol. 21, no. 13, Jul. 2020, doi: 10.3390/ijms21134742.
- [18] E. Chuang, A. M. Hori, C. D. Hesketh, and J. Shorter, "Amyloid assembly and disassembly," *Journal of Cell Science*, vol. 131, no. 8, Apr. 2018, doi: 10.1242/jcs.189928.
- [19] H.-H. (Gavin) Tsai, K. Gunasekaran, and R. Nussinov, "Sequence and Structure Analysis of Parallel β Helices: Implication for Constructing Amyloid Structural Models," *Structure*, vol. 14, no. 6, Jun. 2006, doi: 10.1016/j.str.2006.03.015.
- [20] R. Tycko, K. L. Sciarretta, J. P. R. O. Orgel, and S. C. Meredith, "Evidence for Novel β -Sheet Structures in Iowa Mutant β -Amyloid Fibrils," *Biochemistry*, vol. 48, no. 26, Jul. 2009, doi: 10.1021/bi9002666.
- [21] W. Qiang, W.-M. Yau, Y. Luo, M. P. Mattson, and R. Tycko, "Antiparallel - sheet architecture in Iowa-mutant -amyloid fibrils," *Proceedings of the National Academy of Sciences*, vol. 109, no. 12, Mar. 2012, doi: 10.1073/pnas.1111305109.
- [22] A. K. Buell, "The growth of amyloid fibrils: rates and mechanisms," *Biochemical Journal*, vol. 476, no. 19, Oct. 2019, doi: 10.1042/BCJ20160868.
- [23] T. Uchihara and B. I. Giasson, "Propagation of alpha-synuclein pathology: hypotheses, discoveries, and yet unresolved questions from experimental and

- human brain studies,” *Acta Neuropathologica*, vol. 131, no. 1, Jan. 2016, doi: 10.1007/s00401-015-1485-1.
- [24] S. K. Maji *et al.*, “Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules,” *Science*, vol. 325, no. 5938, Jul. 2009, doi: 10.1126/science.1173155.
 - [25] S. Kroschwald *et al.*, “Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules,” *eLife*, vol. 4, Aug. 2015, doi: 10.7554/eLife.06807.
 - [26] C. P. Brangwynne, P. Tompa, and R. V. Pappu, “Polymer physics of intracellular phase transitions,” *Nature Physics*, vol. 11, no. 11, Nov. 2015, doi: 10.1038/nphys3532.
 - [27] S. Elbaum-Garfinkle *et al.*, “The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 23, Jun. 2015, doi: 10.1073/pnas.1504822112.
 - [28] M. P. Hughes *et al.*, “Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks,” *Science*, vol. 359, no. 6376, Feb. 2018, doi: 10.1126/science.aan6398.
 - [29] J. Wang *et al.*, “A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins,” *Cell*, vol. 174, no. 3, pp. 688-699.e16, Jul. 2018, doi: 10.1016/j.cell.2018.06.006.
 - [30] B. van Treeck, D. S. W. Protter, T. Matheny, A. Khong, C. D. Link, and R. Parker, “RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, Mar. 2018, doi: 10.1073/pnas.1800038115.
 - [31] A. E. Boncella *et al.*, “Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 11, 2020, doi: 10.1073/pnas.1912723117.
 - [32] J. A. Riback *et al.*, “Composition-dependent thermodynamics of intracellular phase separation,” *Nature*, vol. 581, no. 7807, pp. 209–214, May 2020, doi: 10.1038/s41586-020-2256-2.
 - [33] F. Sherman, “[1] Getting started with yeast,” 1991. doi: 10.1016/0076-6879(91)94004-V.
 - [34] C. Baker Brachmann *et al.*, “Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids

for PCR-mediated gene disruption and other applications,” *Yeast*, vol. 14, no. 2, Jan. 1998, doi: 10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

- [35] A. X. Lu, T. Zarin, I. S. Hsu, and A. M. Moses, “YeastSpotter: accurate and parameter-free web segmentation for microscopy images of yeast cells,” *Bioinformatics*, vol. 35, no. 21, Nov. 2019, doi: 10.1093/bioinformatics/btz402.
- [36] E. M. Manders, J. Stap, G. J. Brakenhoff, R. van Driel, and J. A. Aten, “Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy,” *Journal of Cell Science*, vol. 103, no. 3, Nov. 1992, doi: 10.1242/jcs.103.3.857.
- [37] R. J. Planta and W. H. Mager, “The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*,” *Yeast*, vol. 14, no. 5, Mar. 1998, doi: 10.1002/(SICI)1097-0061(19980330)14:5<471::AID-YEA241>3.0.CO;2-U.
- [38] O. M. Subach, P. J. Cranfill, M. W. Davidson, and V. v. Verkhusha, “An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore,” *PLoS ONE*, vol. 6, no. 12, Dec. 2011, doi: 10.1371/journal.pone.0028674.
- [39] Y. O. Chernoff, I. L. Derkach, and S. G. Inge-Vechtomov, “Multicopy SUP35 gene induces de-novo appearance of psi-like factors in the yeast *Saccharomyces cerevisiae*,” *Current Genetics*, vol. 24, no. 3, Sep. 1993, doi: 10.1007/BF00351802.
- [40] L. Xiong, Y. Zeng, R.-Q. Tang, H. S. Alper, F.-W. Bai, and X.-Q. Zhao, “Condition-specific promoter activities in *Saccharomyces cerevisiae*,” *Microbial Cell Factories*, vol. 17, no. 1, Dec. 2018, doi: 10.1186/s12934-018-0899-6.
- [41] P. S. Tsoi, M. D. Quan, K. Choi, K. M. Dao, J. C. Ferreón, and A. C. M. Ferreón, “Electrostatic modulation of hnRNPA1 low-complexity domain liquid–liquid phase separation and aggregation,” *Protein Science*, vol. 30, no. 7, Jul. 2021, doi: 10.1002/pro.4108.
- [42] N. Farahi, T. Lazar, S. J. Wodak, P. Tompa, and R. Pancsa, “Integration of Data from Liquid–Liquid Phase Separation Databases Highlights Concentration and Dosage Sensitivity of LLPS Drivers,” *International Journal of Molecular Sciences*, vol. 22, no. 6, Mar. 2021, doi: 10.3390/ijms22063017.

APPENDICES

Pearson Correlation Coefficient analysis

General overview of the workflow for analysis:

1. Export sets of captures from SlideBook as 16-bit TIFF (OME)
2. Organizing exported image files for analysis
3. Generate image masks with DIC C images to use with fluorescence channel images for each capture (masks are used to delete background/define cell locations within images)
4. Export pixel intensities and image statistics for plotting correlation between fluorescence channels within cells
5. Plot image pixel intensities and export fluorescence channel correlations (Pearson's R, and others) to a spreadsheet
6. Plot correlation scores for each set of captures into a boxplot (or otherwise) for visualization of dataset features across a set of captures, between sets of captures, etc.

IMPORTANT NOTE: For this protocol to work each capture must contain a DIC C image corresponding to two fluorescence channel images (capture the same cells in the same positions in all three images) to calculate correlation between fluorescence channel images.

- For example: One needs a DIC C image plus a 488 channel image (GFP, FITC, etc.) and a 561 channel image (mCherry, RFP, etc.) to calculate correlation between the 488 and the 561 images. On the confocal computer in SlideBook, in the "Capture" window, select "DIC+488+561_ROSS" instead of "Default" for a pre-saved set of imaging parameters that will work for this protocol. Any custom set of imaging parameters only needs to ensure simultaneous capture of DIC C, 488, and 561 channels (or any two fluorescence channels) in order to work.

ANOTHER IMPORTANT NOTE: This protocol is designed to allow the image files to be accessed by the Python scripts from a folder on the desktop. Running the scripts to access folders from the Rdrive, etc. can create issues with file path length. In order to minimize these issues, one should generate a folder on the computer desktop and title it with their name, and they should aim to avoid too much "nesting" of files within folders.

-For example

Ideally the file path length should be short such as:

"Users>ROSSLAB>Desktop>(your name)>Image_analysis> 30C>XYZ_pSup35-sPrLD3", or a title that makes sense for you and your project.

One should avoid excessive file name “nesting” or excessive folder name length such as

“Users>ROSSLAB>Desktop>(your name)>Image_analysis>Rotation_project_Fall_2022>Synthetic_PrLDs_and_Pab1>30C_for_overnight_growth>Day_1>Part_2>Condition_X>...”

-File path length errors can cause the scripts to not run, and concise file paths should help in avoiding these issues.

1. Export sets of captures from SlideBook as 16-bit TIFF (OME)

1. Open the set of captures for export in SlideBook
2. Highlight the captures for export, click on the “Home” tab, click on “→Export”, and click on “16-bit TIFF (OME)”
3. In the “Export Channel Intensities as a 16-bit TIFF” window select either “Selected Image” to export only the highlighted captures, or select “All Images with Same Channels” to export all captures with the same format as the highlighted capture. The “All Images with Same Channels” option can be used to export all 2D images for analysis if the set of captures contains 2D and 3D images mixed together.
4. Next, In the “Output Directory” portion of the “Export Channel Intensities as a 16-bit TIFF” window, make sure “Write Log File” is checked, and click “Browse” to specify the location for SlideBook to export the TIFFs for each capture.
 - Generally, it is best to create an empty folder with a name corresponding to the construct name, condition, etc. for simplicity as each exported TIFF file will not contain the file name seen in SlideBook, only the capture numbers.
5. Once a location for the exported TIFFs has been chosen, click “OK” in the “Export Channel Intensities as a 16-bit TIFF” window.
6. Check to ensure that there are four files generated per capture in the specified folder once the captures are exported. Expect to see the following four files generated for each capture upon successful export:
 1. “Capture #_XY..._Z0_T0_C0” as a Text Document.
 2. “Capture #_XY..._Z0_T0_C0” .TIF or .TIFF
 3. “Capture #_XY..._Z0_T0_C1” .TIF or .TIFF
 4. “Capture #_XY..._Z0_T0_C2” .TIF or .TIFF

-Both TIF or TIFF files will work for analysis

-TIF or TIFF files ending in “C0” should represent DIC C images, “C1” files should represent one fluorescence channel (488/GFP for this protocol), and “C2” files should represent the other fluorescence channel (561/mCherry for this protocol)

2. Organizing exported capture files for analysis

1. Once the capture TIFFs are exported to the new folder (generated in step 1), create a folder within that folder and call it "Text files". Highlight all of the text files for each capture, and move them to the "Text files" folder. (Note: to group text files together within the folder for highlighting all text files at once, right-click and select "Sort by" > "Size". Text files should be much smaller than image files and will be grouped together after sorting.)
2. Next, create three new folders. Each one should contain your initials and a title matching the construct name and condition followed by _488channel, _561channel, or _DICchannel.

Some folder names for example:

XYZ_pSup35-sPrLD2_46C_30min_488channel

XYZ_pSup35-sPrLD2_46C_30min_561channel

XYZ_pSup35-sPrLD2_46C_30min_DICchannel

-Note: In order to run, scripts require that folder names end in _488channel, _561channel, or DICchannel, and folder names should not contain any spaces.

3. Next, move all of the DIC C images (C0 files) to the folder ending in "_DICchannel", move all of the 488 images (C1 files) to the folder ending in "_488channel", and move all of the 561 images (C2 files) to the folder ending in "_561channel".
4. Check to ensure each folder contains the correct number of images, and the right type of images (Only files ending in "_C1" should be in the "_488channel" folder, etc.)

3. Generating image masks with DIC C images for use with fluorescence channel images for each capture

1. After organizing the images into each respective folder, go to the RossLab RSTOR (Rdrive), find the folder titled "Image_analysis_tools", and open this folder.
2. Inside that folder, click on the folder titled "DIC+488+561_quantification", and highlight all of the files within that folder.
3. Copy and paste these 10 items into the folder that contains the folders ending in "_488channel", "_561channel", "_DICchannel", and the folder titled "Text files".
4. This folder should now contain 14 items including the folders for each set of images for each channel, the text files folder, and all 10 items from the folder titled "DIC+488+561_quantification".
5. Right click on the file titled "opts.py" (may only say opts), and select "Edit with NotePad+ +".

6. Once `opts.py` is open, paste the file path corresponding to the input files into the `input_directory` (line 3) of the script. This can be done by manually typing the file path in to NotePad+, or by changing the folder name if the path is mostly consistent between runs.

Example of an acceptable file path input:

```
input_directory =  
"C:/Users/ROSSLAB/Desktop/Your_name/sPrLD2_Pab1_Hyd_Curve/30C/sPrLD2_30C/XYZ_pSup35-sPrLD2_30C_DICchannel"
```

One can paste the file path in from File Explorer to ensure path name consistency, but the slashes will be facing the wrong direction.

Example of incorrect slashes, but correct path:

```
input_directory =  
"C:\Users\ROSSLAB\Desktop\Your_name\sPrLD2_Pab1_Hyd_Curve\30C\sPrLD2_30C\XYZ_pSup35-sPrLD2_30C_DICchannel"
```

-The file path is correct, but the slash type will prevent the script from running. If pasting a file path in from file explorer, ensure slashes face forward.

-If file names vary slightly between runs (e.g. sPrLD2 to sPrLD3), with consistent file naming, one only needs to modify the parts of file paths that change between runs. Changing sPrLD2 to sPrLD3 in NotePad+ will allow the script to run on sPrLD3 files, etc.

7. Next modify the `output_directory` (line 6) to be identical to the `input_directory`, except for the addition of “_OUTPUT” at the end.

Example:

```
output_directory =  
"C:/Users/ROSSLAB/Desktop/Your_name/sPrLD2_Pab1_Hyd_Curve/30C/sPrLD2_30C/XYZ_pSup35-sPrLD2_30C_DICchannel_OUTPUT"
```

8. Hit “Control+S” to save the modified script in your folder before running it.
9. Open Windows PowerShell, type in `cd` (change directory), and navigate to the file path that contains your saved script, then hit enter.

The command line should look somewhat like the following, depending on file path:

PS

```
C:\Users\ROSSLAB\Desktop\Your_name\sPrLD2_Pab1_Hyd_Curve\30C\sPrLD2> _
```

10. Once you have navigated to the correct file location (where your folders containing TIFF images are), in PowerShell, type: `python .\segmentation.py` , and hit enter.
11. You should immediately see the following:
 “Using TensorFlow backend.”, followed by some notifications, and then
 “Preprocessing your images...”, then after each capture is displayed, you should see “Running your images through the neural network...” before the script completes its run.
12. Once the script is done running, check the image folders for a newly generated folder ending in “_DICchannel_OUPUT”. This folder contains your masks, and should have the same number of files as the folder ending in “_DICchannel”.
13. Now, change the name of this folder to end in “_561channel_OUTPUT”.
14. Next, open the image masks folder and change each mask file name from ending in “...Z0_T0_C0”, to end in “...Z0_T0_C2”. This step can be most easily performed by renaming each file separately, changing the last digit only.
15. Once all files are renamed to end in “...Z0_T0_C2”, move on to the next step.

4. Export pixel intensities and image statistics for plotting correlation between fluorescence channels within cells

1. From the image files folder where you have your mask folder ending in “_561channel_OUTPUT”, right-click on
 “CellDimensionsFilter_CreateModifiedImages_FILTER_BORDER_CELLS_QuantifySGs.py”, and select “Edit with NotePad+ +”.
2. In NotePad, navigate to line 28 of the script which should read as follows:
`for (dirname, dirs, files) in os.walk('./XYZ_pSup35-sPrLD2_30C_561channel_OUTPUT/masks').`
3. Change the folder name to reflect the folder that contains your newly generated masks.
4. Hit “Control+S” to save the modified script in your folder before running it.
5. Open Windows PowerShell, type in `cd` (change directory), and navigate to the file path that contains your saved script, then hit enter. If one is running
 “CellDimensionsFilter_CreateModifiedImages_FILTER_BORDER_CELLS_QuantifySGs.py” immediately after running “segmentation.py”, then one only needs to go back to the command line and type: `python .\CellDimensionsFilter_CreateModifiedImages_FILTER_BORDER_CELLS_QuantifySGs.py` and hit enter since both saved scripts are in the same location (file path).
6. You should see the script running after hitting enter, and new files should be generated in the same folder as the script/image folders indicating the script is running successfully.
7. Once the script is done running, open the spreadsheet titled
 “Processed_Image_Data_2p0_multiplier” in that same folder. You should see a list of values for each capture representing the channels for each cell within

each capture. If the spreadsheet is empty, try checking that image mask file names are all changed from C0 to C2, and that the folder name/file path to the image masks is typed in to NotePad+ + correctly.

-Note: This script will always generate a spreadsheet with the same name (Processed_Image_Data_2p0_multiplier). In order to open more than one of these spreadsheets at a time, or to save them in the same folder, they need to be renamed to have a unique title that makes sense for your project. These titles will be fed in to later scripts as file names so consistency may help with organization.

5. Plot image pixel intensities and export fluorescence channel correlations to a spreadsheet

1. In the same folder where your newly generated spreadsheet is, right-click on "plot_intensity_correlation_ScatterPlots.py" and select "Edit with NotePad+ +"
2. Then scroll down to line 36 which should read as follows: `file_path = './XYZ_pSup35-sPrLD2_30C_488channel/'`, and change the file path to reflect the name of the folder that contains your 488channel images (C1 files).
3. Next, scroll down to line 59 which should read as follows: `file_path = './XYZ_pSup35-sPrLD2_30C_561channel/'`, and change the file path to reflect the name of the folder that contains your 561channel images. (C2 files).
4. Hit "Control+S" to save the modified script in your folder before running it.
5. Open Windows PowerShell, type in `cd` (change directory), and navigate to the file path that contains your saved script, then hit enter. If one is running "plot_intensity_correlation_ScatterPlots.py" immediately after running "CellDimensionsFilter_CreateModifiedImages_FILTER_BORDER_CELLS_Q_uantifySGs.py", then one only needs to go back to the command line and type: `python .\plot_intensity_correlation_ScatterPlots.py` and hit enter since both saved scripts are in the same location (file path).
6. You should see the script running after hitting enter, and new files should be generated in the same folder as the script/image folders indicating the script is running successfully.
7. Once the script is done running, open the spreadsheet titled "PearsonCorrelationCoefficients_Results_with_tau_rho" in that same folder. You should see lists of values for each capture representing each cell within each capture. If the spreadsheet is empty, try checking if the file name for the image folders is typed in to NotePad+ + correctly, and that no other changes to the script were made.

6. Plot correlation scores for each set of captures into a boxplot (or otherwise) for visualization of dataset features.

1. In the same folder where your newly generated spreadsheet (PearsonCorrelationCoefficients...) is, right-click on "boxplot_CorrelationCoefficients.py", and select "Edit with NotePad+ +"

2. Navigate to line 12, which should read as follows:
`files = ['PearsonCorrelationCoefficients_Results_with_tau_rho.csv']`.

-Note: If you have not modified the name of your new spreadsheet generated in step 5, the script is ready to run with no modification needed. However, if you wish to plot multiple boxplots in the same figure, simply add a comma between the file name in quotations to plot multiple spreadsheets.

Example:

```
files = ['PearsonCorrelationCoefficients_Results_with_tau_rho.csv']  
becomes  
files = ['Ex1_PearsonCorrelationCoefficients_Results_with_tau_rho.csv',  
'Ex2_PearsonCorrelationCoefficients_Results_with_tau_rho.csv' ]
```

One can plot several boxplots together as long as each file name is in quotations, and separated by a comma like what is shown above. All files in the list must be surrounded by the brackets, and each spreadsheet must have a unique name.

3. Hit “Control+S” to save the modified script in your folder before running it (if applicable).
4. Open Windows PowerShell, type in `cd` (change directory), and navigate to the file path that contains your saved script, then hit enter. If one is running “`boxplot_CorrelationCoefficients.py`” immediately after running “`plot_intensity_correlation_Scatterplots.py`”, then one only needs to go back to the command line and type: `python .\ boxplot_CorrelationCoefficients.py` and hit enter since both saved scripts are in the same location (file path).
5. You should see the script running after hitting enter, and a new file should be generated in the same folder as the script/image folders indicating the script ran successfully.
6. This new file contains your boxplot(s). Open this file, and reveal your new data!

Note: The instructions for step 6 also apply to the script titled “`violinplot_CorrelationCoefficients.py`”, as they both accept the same file type input (`'PearsonCorrelationCoefficients_Results_with_tau_rho.csv'`).

Partition Coefficient analysis

General overview of the workflow for analysis:

1. Export captures from SlideBook as 16-bit TIFFs (OME)
2. Organizing exported image files for analysis
3. Generate image masks with DIC C images to use with fluorescence channel images for each capture (masks are used to delete background/define cell locations within images)
4. Subtract “blank” or “background” image from images of cells
5. Filter vacuole & nuclear pixels out from pixel set making up each cell, define cytoplasm & stress granule pixels in 3 color image stacks
6. Export pixel intensity data for 488 channel image pixels (protein of interest channel)
7. Calculate partition coefficient from the pixel intensities for each cell in the set of images
8. Plot partition coefficients

IMPORTANT NOTE: For this protocol to work each capture must contain a DIC C image corresponding to three fluorescence channel images (capture the same cells in the same positions in all four images) to calculate partition coefficient between fluorescence channel images.

- For example: One needs a DIC C image plus a 405 channel image (BFP), a 488 channel image (GFP, FITC, etc.) and a 561 channel image (mCherry, RFP, etc.) to calculate partitioning of the 488 channel images. On the confocal computer in SlideBook, in the “Capture” window, select “DIC+405+488+561_ROSS” instead of “Default” for a pre-saved set of imaging parameters that will work for this protocol. Any custom set of imaging parameters only needs to ensure simultaneous capture of DIC C, 405, 488, and 561 channels (or any three fluorescence channels) in order to work.

ANOTHER IMPORTANT NOTE: This protocol is designed to allow the image files to be accessed by the Python scripts from a folder on the desktop. Running the scripts to access folders from the Rdrive, etc. can create issues with file path length. In order to minimize these issues, one should generate a folder on the computer desktop and title it with their name, and they should aim to avoid too much “nesting” of files within folders.

-For example

Ideally the file path length should be short such as:

“Users>ROSSLAB>Desktop>(your name)>Image_analysis> 30C>XYZ_pSup35-sPrLD3”,

or a title that makes sense for you and your project.

One should avoid excessive file name “nesting” or excessive folder name length such as

“Users>ROSSLAB>Desktop>(your name)>Image_analysis>Rotation_project_Fall_2022>Synthetic_PrLDs_and_Pab1>30C_for_overnight_growth>Day_1>Part_2>Condition_X>...”

-File path length errors can cause the scripts to not run, and concise file paths should help in avoiding these issues.

1. Export sets of captures from SlideBook as 16-bit TIFF (OME)

1. Open the set of captures for export in SlideBook
2. Highlight the captures for export, click on the “Home” tab, click on “→Export”, and click on “16-bit TIFF (OME)”
3. In the “Export Channel Intensities as a 16-bit TIFF” window select either “Selected Image” to export only the highlighted captures, or select “All Images with Same Channels” to export all captures with the same format as the highlighted capture. The “All Images with Same Channels” option can be used to export all 2D images for analysis if the set of captures contains 2D and 3D images mixed together.
4. Next, In the “Output Directory” portion of the “Export Channel Intensities as a 16-bit TIFF” window, make sure “Write Log File” is checked, and click “Browse” to specify the location for SlideBook to export the TIFFs for each capture.
 - Generally, it is best to create an empty folder with a name corresponding to the construct name, condition, etc. for simplicity as each exported TIFF file will not contain the file name seen in SlideBook, only the capture numbers.
5. Once a location for the exported TIFFs has been chosen, click “OK” in the “Export Channel Intensities as a 16-bit TIFF” window.
6. Check to ensure that there are four files generated per capture in the specified folder once the captures are exported. Expect to see the following four files generated for each capture upon successful export:
 1. “Capture #_XY..._Z0_T0_C0” as a Text Document.
 2. “Capture #_XY..._Z0_T0_C0” .TIF or .TIFF
 3. “Capture #_XY..._Z0_T0_C1” .TIF or .TIFF
 4. “Capture #_XY..._Z0_T0_C2” .TIF or .TIFF
 5. “Capture #_XY..._Z0_T0_C3” .TIF or .TIFF

-Both TIF or TIFF files should work for analysis

-TIF or TIFF files ending in “C0” should represent DIC C images, “C1” files should represent one fluorescence channel (488/GFP for this protocol), “C2” files should represent another channel (561/mCherry for this protocol), “C3” files should represent the third channel (405/mTagBFP2 for this protocol)

2. Organizing exported capture files for analysis

1. Once the capture TIFFs are exported to the new folder (generated in step 1), create a folder within that folder and call it "Text files". Highlight all of the text files for each capture, and move them to the "Text files" folder.
2. Next, create three new folders. Each one should contain your initials and a title matching the construct name and condition followed by, _405channel, _488channel, _561channel, or _DICchannel.

Some folder names for example:

XYZ_pSup35-sPrLD2_46C_30min_405channel
XYZ_pSup35-sPrLD2_46C_30min_488channel
XYZ_pSup35-sPrLD2_46C_30min_561channel
XYZ_pSup35-sPrLD2_46C_30min_DICchannel

-Note: In order to run, scripts require that folder names end in _405channel, _488channel, _561channel, or _DICchannel, and folder names should not contain any spaces.

3. Next, move all of the DIC C images (C0 files) to the folder ending in "_DICchannel", move all of the 488 images (C1 files) to the folder ending in "_488channel", move all of the 561 images (C2 files) to the folder ending in "_561channel", and move all of the 405 images (C3 files) to the folder ending in "_405channel".
4. Check to ensure each folder contains the correct number of images, and the right type of images (Only files ending in "_C1" should be in the "_488channel" folder, etc.)

3. Generating image masks from DIC C images for use with fluorescence channel images for each capture

1. After organizing the images into each respective folder, go to the RossLab RSTOR (Rdrive), find the folder titled "Image_analysis_tools", and open this folder.
2. Inside that folder, click on the folder titled "Partition_coefficient" and highlight all of the files within that folder, except the word document.
3. Copy and paste the 7 python scripts & 2 folders into the folder that contains the folders ending in "_405channel", "_488channel", "_561channel", "_DICchannel", and the folder titled "Text files".
4. This folder should now contain the folders for each set of images for each channel, the text files folder, and the items the folder titled "Partition_coefficient".
5. Right click on the file titled "opts.py" (may only say opts), and select "Edit with NotePad+ +".
6. Once opts.py is open, paste the file path corresponding to the input files into the input_directory (line 3) of the script. This can be done by manually typing

the file path in to NotePad+ +, or by changing the folder name if the path is mostly consistent between runs.

Example of an acceptable file path input:

input_directory =

"C:/Users/ROSSLAB/Desktop/Your_name/sPrLD2_Pab1_Hyd_Curve/30C/sPrLD2_30C/ XYZ_pSup35-sPrLD2_30C_DICchannel"

One can paste the file path in from File Explorer to ensure path name consistency, but the slashes will be facing the wrong direction.

Example of incorrect slashes, but correct path:

input_directory =

"C:\Users\ROSSLAB\Desktop\Your_name\sPrLD2_Pab1_Hyd_Curve\30C\sPrLD2_30C\ XYZ_pSup35-sPrLD2_30C_DICchannel"

-The file path is correct, but the slash type will prevent the script from running. If pasting a file path in from file explorer, ensure slashes face forward.

-If file names vary slightly between runs (e.g. sPrLD2 to sPrLD3), with consistent file naming, one only needs to modify the parts of file paths that change between runs. Changing sPrLD2 to sPrLD3 in NotePad+ + will allow the script to run on sPrLD3 files, etc.

7. Next modify the output_directory (line 6) to be identical to the input_directory, except for the addition of "_OUTPUT" at the end.

Example:

output_directory =

"C:/Users/ROSSLAB/Desktop/Your_name/sPrLD2_Pab1_Hyd_Curve/30C/sPrLD2_30C/XYZ_pSup35-sPrLD2_30C_DICchannel_OUTPUT"

8. Hit "Control+S" to save the modified script in your folder before running it.
9. Open Windows PowerShell, type in cd (change directory), and navigate to the file path that contains your saved script, then hit enter.

The command line should look somewhat like the following, depending on file path:

PS

C:\Users\ROSSLAB\Desktop\Your_name\sPrLD2_Pab1_Hyd_Curve\30C\sPrLD2> _

10. Once you have navigated to the correct file location (where your folders containing TIFF images are), in PowerShell, type: python .\segmentation.py , and hit enter.
11. You should immediately see the following:

- “Using TensorFlow backend.”, followed by some notifications, and then “Preprocessing your images...”, then after each capture is displayed, you should see “Running your images through the neural network...” before the script completes its run.
12. Once the script is done running, check the image folders for a newly generated folder ending in “_DICchannel_OUTPUT”. This folder contains your masks, and should have the same number of files as the folder ending in “_DICchannel”.
- 4. Subtracting a “blank” or “background” image from images of cells**
1. From the image files folder where you have your mask folder ending in “_DICchannel_OUTPUT”, right-click on “Background_fluorescence_subtraction.py”, and select “Edit with NotePad+”.
 2. In NotePad, navigate to line 13 of the script which should read as follows:
Bg_images = [“./sPrLD...488channel/Capture...C1.tif”], [“./sPrLD...C2.tif”], [“./sPrLD...C3.tif”]
 3. Change the image file names in line 13 to reflect the exact name of the background images (C1.tif for 488 image, C2.tif for 561 image, etc.)
 4. Hit “Control+S” to save the modified script in your folder before running it.
 5. Open Windows PowerShell, type in cd (change directory), and navigate to the file path that contains your saved script, then hit enter. If one is running “Background_fluorescence_subtraction.py” immediately after running “segmentation.py”, then one only needs to go back to the command line and type: python .\ Background_fluorescence_subtraction.py and hit enter since both saved scripts are in the same location (file path).
 6. You should see the script running after hitting enter, and new images should be generated in the same folder as the script/image folders indicating the script is running successfully.
 7. Once the script is done running, create a new folder called “raw images”, select the 405, 488, and 561 image folders, and move these folders to the “raw images” folder.
 8. Create three empty folders with the same names as the three folders moved to “raw images” (the folders ending in _405channel, _488channel, etc.), then put the new images created by background subtraction into the appropriate folders. C1 images in 488channel folder, C2 images in 561channel, etc.
- 5. Filter vacuole & nuclear pixels out from pixel set making up each cell, define cytoplasm & stress granule pixels in 3 color image stacks**
1. In the same folder where your newly generated image folders containing the background subtracted images are, right-click on “QuantSG.py” and select “Edit with NotePad+”
 2. Then scroll down to line 24 which should read as follows: for (dirname, dirs, files) in os.walk('./sPrLD3m6...OUTPUT/masks'): , and change the file path to reflect the name of the folder that contains your DICchannel_OUTPUT masks (C0 files).

3. Hit "Control+S" to save the modified script in your folder before running it.
4. Open Windows PowerShell, type in cd (change directory), then navigate to the file path that contains your saved script, and hit tab to complete the file path. Once you see "QuantSG.py" in powershell, type "-e C1.tif -c 488channel -e_sg C2.tif -c_sg 561channel -e_cyt C3.tif -c_cyt 405channel". These command line arguments are crucial to enter for the script to run. If one is running "QuantSG.py" immediately after running "Background_fluorescence_subtraction.py", one needs to go back to the command line and type: python .\ QuantSG.py "-e C1.tif -c 488channel -e_sg C2.tif -c_sg 561channel -e_cyt C3.tif -c_cyt 405channel" then hit enter
5. You may see the script running after hitting enter, and new files may be generated in the same folder as the script/image folders indicating the script is running successfully.
6. Once the script is done running, you should see a spreadsheet created titled "Processed_image_data...multiplier.csv"

6. Export pixel intensity data for 488 channel image pixels (protein of interest channel)

1. In the same folder where your newly generated spreadsheet (Processed_image_data...) is, right-click on "get_POI_data.py", and select "Edit with NotePad+ +"
2. Navigate to line 27, which should read as follows:
dir_leader = './sPrLD3m6_46C_30min_'
3. Change line 27 to reflect the beginning of the folder name of the folder that contains your 488channel images (C1 files).
Example: The folder "sPrLD3m6_46C_30min_488channel" is entered correctly in line 27 when line 27 contains "sPrLD3m6_46C_30min_" and not 488channel at the end. The folder name is broken up by this script at the underscores in the title. Include the entire folder title except for 488channel.
4. While this script is running it will actively filter cells out due to low expression or overexposure with the lasers while imaging. Cells that are filtered out will be displayed in the command line as the script is running. Each filtered cell will display the cell number, the median GFP intensity of the cell, and whether it was above or below median intensity thresholds used for filtering.
Note: There are three GFP intensity filters in get_POI_data.py
 1. Overexposure filter: cell median exceeding 50,000 (on a scale from 0 to 65,535)
 2. Low expression filter (sliding): cell median not exceeding 10% of max GFP intensity for a given image
 3. Low expression filter (fixed): cell median not exceeding 1000 in cases where 10% of image max is less than 1000
5. Hit "Control+S" to save the modified script in your folder before running it.

7. Calculate partition coefficient from the pixel intensities for each cell in the set of images

1. In the same folder as before, right-click on “calculate_PartitionCoefficient_3colorImaging. py”, and select “Edit with NotePad+ +”
2. Navigate to line 15, which should read as follows:
dir_leader = './sPrLD3m6_46C_30min_ '
3. Change line 15 to reflect the beginning of the folder name of the folder that contains your 488channel images (C1 files).
Example: The folder “sPrLD3m6_46C_30min_488channel” is entered correctly in line 27 when line 27 contains “sPrLD3m6_46C_30min_” and not 488channel at the end. The folder name is broken up by this script at the underscores in the title. Include the entire folder title except for 488channel.
4. Hit “Control+S” to save the modified script in your folder before running it.
5. Open Windows PowerShell, navigate to the file path that contains your saved script, then hit enter. If one is running “calculate_PartitionCoefficient_3colorImaging.py” immediately after running “get_POI_data.py”, then one only needs to go back to the command line and type: python .\ calculate_PartitionCoefficient_3colorImaging.py and hit enter since both saved scripts are in the same location (file path).
6. You should see the script running after hitting enter, and a spreadsheet containing partition coefficients will be generated in the same folder as the script/image folders indicating the script ran successfully.

8. Plot partition coefficients

1. In the same folder where your newly generated spreadsheet is, right-click on “plot_PartitionCoefficients.py”, and select “Edit with NotePad+ +”
2. The file name is broken up into 3 parts. The file leader (line 7) will need to be identical in each spreadsheet one wishes to plot. The file ending (line 8) will also need to be identical in each spreadsheet one wishes to plot. The constructs in single quotation marks (line 9), will be the unique portion of each spreadsheet name. This unique name will be the label for the data on the X axis. Choose construct names you wish to see on the final plot. Make sure each construct name is in quotes, and separated by a comma.
3. To change the X-axis, navigate to line 40, and change the quoted text from 'Construct', to a title that makes sense for your data. Ensure it is in single quotes just as the temporary title is.
4. Hit “Control+S” to save the modified script in your folder before running it.
5. Open Windows PowerShell, navigate to the file path that contains your saved script, then hit enter. If one is running “plot_PartitionCoefficients.py” immediately after running “calculate_PartitionCoefficient_3colorImaging.py”, then one only needs to go back to the command line and type: python .\ plot_PartitionCoefficients.py and hit enter since both saved scripts are in the same location (file path).
6. You should see the script running after hitting enter, and a plot containing partition coefficients will be generated in the same folder as the script/image folders indicating the script ran successfully.