THE UNIVERSITY OF CHICAGO

INVESTIGATION OF THE MOLECULAR MECHANISM OF THE STRESS-TRIGGERED CONDENSATION OF POLY(A)-BINDING PROTEIN

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

RUOFAN CHEN

CHICAGO, ILLINOIS

AUGUST 2022

For my husband Jinzhi

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Everyone who I have worked with during the past five years has been incredibly kind and helpful. I would like to first express my gratitude to my advisor Professor Tobin R. Sosnick for sharing his rich knowledge and insights and providing endless support along the way. His enthusiasm for science, dedication to search for the truth, and willingness to attend to every detail of research on his own despite how busy he is have been inspiring me to strive for my best even during discouragement. I would also like to thank Professor D. Allan Drummond who provided incredible help with this project, for teaching me how to think broadly when designing experiments, analyze the data rigorously and critically, and present the data clearly and compellingly. This work is only possible with the help and guidance from Tobin and Allan. Besides how to approach scientific problems, I am also inspired by how they think and work as a scientist, and how they coordinate with and guide different people including their students. I would also like to thank Professor Allison H. Squires and Professor Juan L. Mendoza for being on my thesis committee.

I would like to thank all the members from Sosnick and Drummond lab for all the discussions, valuable feedback and kind help. I would like to thank Darren Kahan and Julia Shangguan who I worked with on this project for what I learned from them and their stellar contribution to this work. I would like to thank Dr. Haneul Yoo and Dr. Jared Bard for providing extensive help and insightful perspectives, and Dr. Joshua Riback who set the foundation of this project. I want to thank Dr. Xiangda Peng for his help with but not limited to computational studies and Dr. Adam Maciek Zmyslowski who taught me basically every technique used in the lab. I would like to thank Isabelle Gagnon who takes care of almost everything to make the lab a best place to work in. I would also like to thank Joseph Sachleben (Biomolecular NMR Core Facility) and Elena Solomaha (Biophysics Core Facility) for their help.

I would like to thank my friends for their support and being a best part of my life out of science. I would like to thank my parents Huoying Yao and Wenhua Chen for their love and

support for me to pursue anything that I love since my childhood, which is not something I should take for granted. I would like to thank my husband Jinzhi Lu who encouraged, supported and inspired me throughout my Ph.D.

ABSTRACT

Poly(A)-binding protein (PABP, Pab1 in budding yeast), is a canonical stress granule marker that is consistently recruited to stress granules under a range of stresses. Pab1 autonomously condenses upon heat shock or starvation-induced acidification both *in vivo* and *in vitro*, and Pab1's condensation has been shown to be an adaptive stress response which promotes cellular fitness during stressed conditions.

It was unclear how Pab1 senses thermal stress or acidification, transduces either stress and condenses. Particularly, Pab1's RRMs are necessary and sufficient for condensation whereas the IDR P domain is not required, which is opposite to the common observation that IDR is essential for condensation. Pab1's condensation with an LCST also contrasts most condensing proteins with a UCST. The aim of this thesis is to investigate the molecular mechanism of Pab1's condensation through a combination of different biophysical measurements. An outstanding issue of the field is that there is no readily available approach to probe the condensates with high structural resolution, due to that the usual set of high resolution methods are not applicable to condensates. We address this problem by innovatively applying HDX-MS to Pab1 condensates to probe the hydrogen bonding in the condensates.

In Chapter 2, I present HDX-MS data of both monomeric Pab1 and Pab1 condensates, which reveals that RRMs partially unfold upon condensation. RRMs exhibit different levels and patterns of unfolding with RRM3 remaining largely folded. HDX data of condensates exhibits high level of heterogeneity, which points to a structural diversity and interaction heterogeneity in the condensates. HDX-MS also indicates that pH- and Temperature-induced condensates are structurally similar. Our HDX data uncovers that partial unfolding underlies Pab1's condensation process, which supports the view that disorder and quinary interactions underlie the condensation of Pab1 as they do in other systems that condense, just that partial unfolding is an early step of the process.

In Chapter 3, I show that RRMs participate differently in condensation. RRMs have different activation temperatures above which they partially unfold and participate in con-

densation. From this observation we porposed a "sequential activation" model. Co-demixing experiments indicate that the activation is required for an RRM to be engaged in condensation. We term the mechanism "thermodynamic specificity" that an activated RRM will strongly interact only with other activated RRMs rather than inactivated ones.

In Chapter 4, I discuss other molecular factors of Pab1's condensation, including histidines' titration, nucleation effect, intramolecular inhibition, and linkers' flexibility. Importantly, we found that histidines partly confer Pab1's sensitivity to pH. Furthermore, SAXS studies show that Pab1's linkers are flexible, and the condensation is not auto-inhibited. FRET further confirms that Pab1 exhibits no significant large-scale conformational change upon condensation. Lastly, I present data from simulation for testing the mode of RRM interactions, which suggests non-specific interactions between RRMs.

In Chapter 5, I summarize the findings and conclusions described in previous chapters and address some open questions and future directions. Overall, our multi-facet study investigates the molecular mechanism of Pab1's stress-triggered condensation and highlights the potential of HDX-MS as a novel high-resolution analysis tool for biomolecular condensates.

CHAPTER 1

INTRODUCTION

In this chapter, I begin with a general introduction of biomolecular condensates, and the physical model through which the field establishes the understanding of the principles of biomolecular condensation. Stress granule is a well-studied class of biomolecular condensates which forms as intracellular assemblies under stressed conditions including heat shock, oxidative stress, starvation, UV, and perturbative chemicals. Pab1 (Poly(A)-binding protein in yeast), a canonical marker and a key protein component of stress granules, is consistently recruited to stress granules under a range of stresses. Previous study shows that Pab1 undergoes stress-triggered condensation with high sensitivity to stress which has been shown to be an adaptive stress response.

1.0.1 Biomolecular condensates

Biochemical reactions in the cells are complicated and they are implicated in different biological processes to maintain the cellular fitness and functions. A key question is how cells regulate a complex collection of biochemical reactions, both spatially and temporally. There are thousands of biomolecules in the tight space of cells, with a total cytosolic protein concentrations estimated at roughly 3 mM [19]. Cells need to tightly control the cellular biomolecules for them to play diverse roles in different biochemical reactions without the formation of deleterious protein aggregates.

Biomolecular condensates are intracellular compartments without a surrounding membrane that function to concentrate proteins and nucleic acids [5]. They are also termed membraneless organelles, to distinguish them from the traditional membrane-bound organelles including the mitochondria and endoplasmic reticulum. Membraneless organelles have been studied since over 100 years ago. While some dynamic or semi-fluid properties were discovered [56], it remained unclear what is the underlying physical mechanism. In

2009, Brangwynne et al. showed that P granules in *C. elegans* exhibit liquid-like behaviors including fusion, dripping, wetting events, and spherical appearance [7]. These properties are hallmarks of phase separation, a process where a homogeneous mixture divides itself into two phases - a dense and a dilute phase. The thermodynamics underlying phase separation is that the enthalpy promotes demixing contributed by favorable interactions within each phase and overcomes the unfavorable entropy loss associated with the concentrated phase (Figure 1.1A-C). The phase diagram often exhibits two types of boundaries, UCST (upper critical solution temperature) and LCST (lower critical solution temperature) (Figure 1.1D). The mixture components are miscible below the critical temperature for all compositions in an LCST case or above the critical temperature in a UCST case.

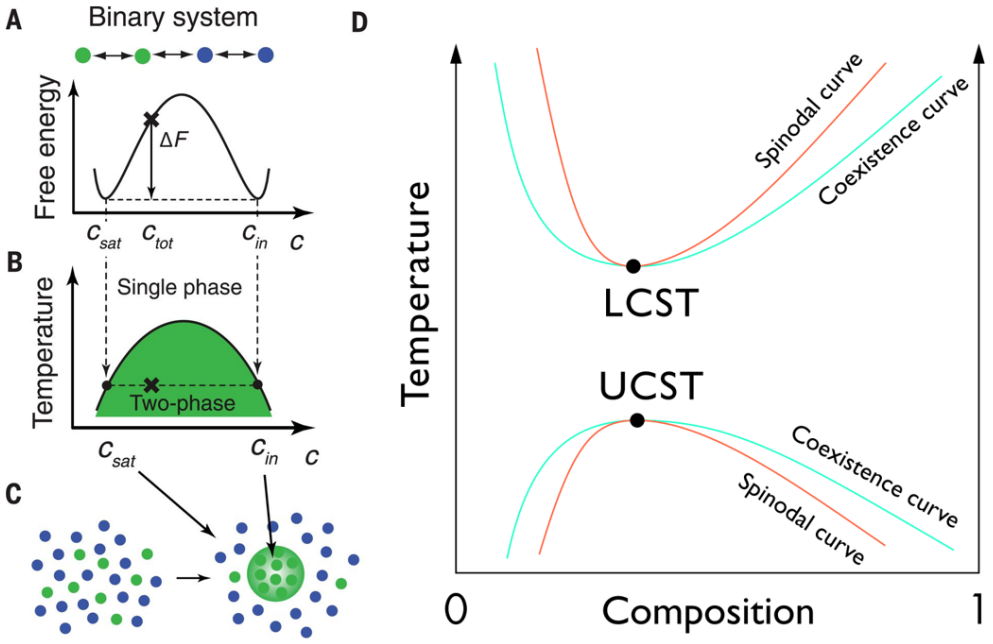


Figure 1.1: Thermodynamics of phase separation. (A) A mixed binary system demixes into two phases to lower its overall free energy. (B) The compositions of two phases are determined by the temperature (more precisely, the free energy construction at different temperatures). (C) The dense phase and the dilute phase. (D) the phase diagram under LCST or USCT regime. The coexistence curve is also called the binodal curve. Figure(A)-(C) is from [60] and (D) is from Wikipedia.

A study on nucleoli from *Xenopus laevis* revealed similar coalescing liquid properties [8]. Since then, a growing number of studies have revealed that cells form biomolecular condensates via phase separation and utilize them to organize cellular biochemistry [55]. Biomolecular condensates can selectively sequester biomolecules into a compartment and promote biological reactions, and maintain a stabilized cellular concentration of certain types of molecules, buffering environmental noise [18]. There is a variety of cellular biomolecular condensates, including cytoplasmic ones (stress granule, P-body, germ granule) and nuclear ones (Cajal body, nucleolus), carrying out various functions in different contexts (Figure 1.2).

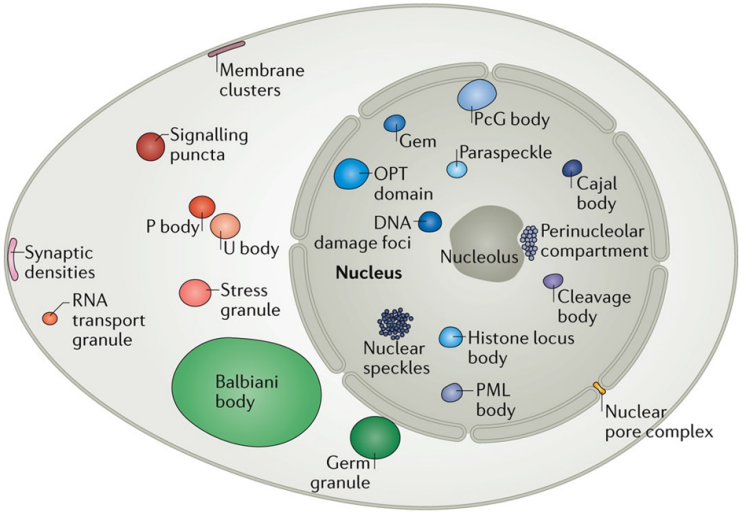


Figure 1.2: Biomolecular condensates/membraneless organelles in the cells. Figure is from [5].

Biomolecular condensates are tightly regulated inside the cells in terms of formation, droplet size, composition, and material state through a range of mechanism including post-translational modifications [53, 68], inclusion of RNA [61, 58], and other mechanism yet well-understood. Condensates can be affected by pathological mutations and other factors which may lead to aberrant aggregates formation [55].

It is important to point out the difference between biomolecular condensates and protein aggregates. While both of them are the assemblies of proteins (or/and nucleic acids), protein

aggregates are usually used to refer to a non-native state resulting from misfolding events in pathological conditions and are irreversible in general cases. Although not all bimolecular condensates have a clearly defined function, an increasing number of studies have identified biomolecular condensates' roles in various contexts. For example, P-body regulates mRNA decay, silencing, and storage [44, 23], and germline P granules regulate the cell cycle [62]. Besides that, bimolecular condensates are generally dynamic and reversible so that cells can dissolve and reuse the component molecules [35, 33].

The field has used the concept of "phase separation" to describe biomolecular condensates as there is a tradition of approximating proteins as biopolymers in order to apply polymer physics theory to proteins. Biomolecular condensates typically are considered to form through liquid-liquid phase separation (LLPS). There are issues with the generic use of the term LLPS to describe all condensates including:

1. Biomolecular condensates span a range of material states including liquid [17], hydrogel [57] and solid [55]. LLPS implies a liquid form which is not necessarily the case. In some cases, bimolecular condensates can have a second hardening/maturation process after initial LLPS, where the condensates convert from initial fluid to a more solid-like, less dynamic state [43].

2. In the Flory-Huggins model [22, 28] of polymer solution theory, there are some approximations including mean-field approximation of enthalpy and neglecting the effect on entropy from the connection between monomer units. Those approximations can have significant deviations for protein in some cases. For example, while Flory-Huggins model implies a constant concentration of both the dilute and dense phases at a given temperature and a reversible process, those are not always observed for bimolecular condensates. In addition, structure formation can take place after the initial phase transition in the condensed phase due to increased local concentration. Examples include amyloid formation [33] and maturation of RNP granules [43].

The term of bimolecular condensates addresses those issues as it is a more general concept

without specifying the mechanism or the material state unequivocally. In this thesis, I will refer to the process of the formation of biomolecular condensates using the term "condensation" instead of "phase separation". Additionally, while RNA has been identified to be a key component and an important regulatory determinant for many condensation processes, the emphasis of this thesis will be on protein condensation.

1.0.2 The formation and interactions of biomolecular condensates

The general view is that protein condensation is driven by weak, multivalent interactions which form a non-stoichiometric and dynamic network, inheriting the classic concept of multivalency from polymer science. Multivalent interactions can be mediated by folded domains in multi-domain constructs (e.g., engineered constructs containing repeats of SH3/PRM or SUMO/SIM undergo phase transition [42, 6]), or intrinsically disordered protein/region (IDP/IDR) with multiple interacting motifs (short linear motif, SLiM) along the sequence [25]. IDP/IDR is a protein (region) that lacks a fixed three-dimensional structure and usually adopts a broad conformational ensemble [15]. The sequence of IDR is frequently found to be low complexity and have a biased composition of amino acid residues [33]. On the other hand, low complexity regions (LCRs) in protein are often unstructured, so there is an intersection between LCR and IDR. Though LCR and IDR are different concepts, as some LCR can form secondary structures, and IDR sequence is not necessarily low complexity, LCR and IDR overlap a lot in the context of condensation, and the field often uses the two terms interchangeably.

IDR are typically enriched in polar amino acids and sometimes charged residues, and are depleted of hydrophobic residues with aliphatic side chains [14]. Aromatic residues are also frequently found in IDRs. The sequence feature of a biased composition endows IDR to mediate multivalent interactions. As a result, condensing proteins from reported studies are highly enriched in IDRs, and their IDRs has been shown to be required for condensation in hnRNPA1 [51], FUS [55], LAF-1 [17] and Sup35 [24]. Some IDRs can condense alone

without any folded domain in the structure [53].

Protein condensation can be triggered by different conditions including low salt concentration [17, 24], low temperature [51] (or high temperature [57] in some cases), or low pH [57, 40, 24]. The major interest of the field is to understand how the protein's primary sequence encodes different condensation behavior. Studies have identified a number of molecular factors, often highlighting the roles of aromatic residues and charged residues [53] which mediate pi-pi stacking interactions and cation-pi interactions as important driving forces for condensation. Specifically, the numbers of Tyrosine and Arginine have been found to govern the concentration threshold for condensate formation in the FUS family of proteins [67]. Polar residues can also contribute via dipolar interactions. The patterning of these residues also impacts the condensing behavior. For example, clustering of charges residues to form blocks of different charges is favored compared to a uniform distribution of charges [53, 54]. The sequence composition and patterning can both affect the type of phase behavior which is an interplay between entropy and enthalpy. For example, the inclusion of hydrophobic residues contributes to LCST phase behavior which is entropically driven, whereas polar residues and oppositely charged residues promotes UCST phase behavior which is enthalpically driven [47]. Residues including glycine, serine and glutamine can tune the material state of condensed FUS droplets, where glycine promotes liquidity while serine and glutamine promote hardening [67].

Based on the framework of Flory-Huggins theory and more recent work that identified sequence-encoded information of condensation, stickers-and-spacers model [11] was developed to formalize the phase behavior of those associative biopolymers. Residues/motifs that mediate or participate in attractive interactions are considered as stickers while other parts of the protein sequence interspersed between stickers are considered as spacers (Figure 1.3). The stickers-and-spacers model is generalizable as it has minimal if any restrictions on the identity of stickers and spacers. Stickers can be interaction patches on a folded protein, or folded binding domains connected by linkers as spacers on a multivalent construct, or

certain types of residues/SLiMs on an IDR. Three types of interactions including sticker-sticker, sticker-spacer, and spacer-spacer interactions with different strengths are included in the model. The relative strength of the three interactions together with the constructs' valency of stickers determine the percolation concentration threshold, resulting from an equilibrium between interchain and intrachain contacts.

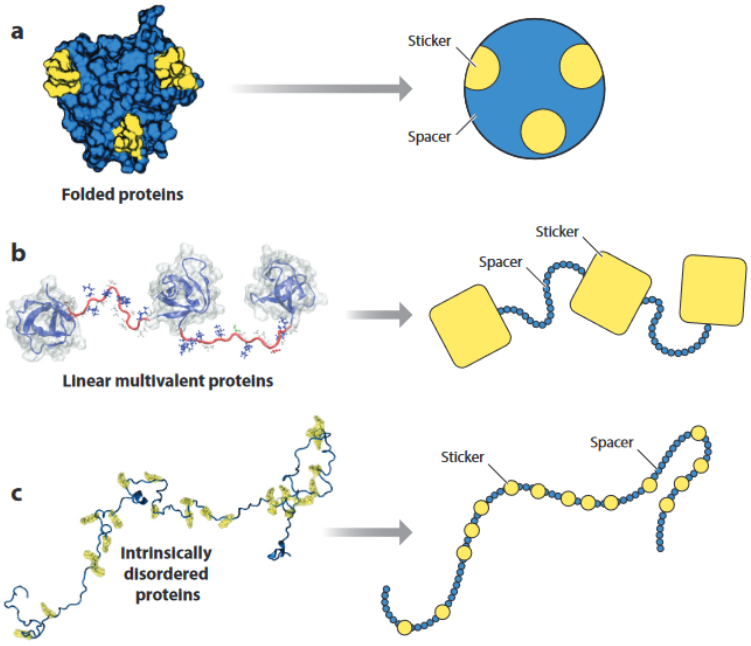


Figure 1.3: Schematic of stickers and spacers in different protein systems including folded protein, multi-modular proteins connected by linkers and IDP. Figure is from [11].

1.0.3 Stress granules

Stress granules are assemblies of ribonucleoproteins (RNP) that form in response to changes in the cellular environment including heat and oxidative stress, starvation, UV, and perturbative chemicals [30, 10, 40]. Stress granule is one of cytoplasmic biomolecular condensates, and has been shown to be dynamic from FRAP experiment (Fluorescence recovery after photobleaching [3], a method for determining the diffusion kinetics) [35] but some study suggested that stress granule has a less dynamic core [63, 30].

Experimental evidence suggests that stress granules are assemblies of mRNP (messenger ribonucleoprotein) that form from mRNA stalled in translation initiation when translation initiation is inhibited by chemicals or other stresses [49]. Stress granules are composed of mRNA, RNA-binding proteins including PABP (poly(A)-binding protein) and translation initiation factors (eIF2, eIF3, eIF4E, eIF4G), and other non-RNA-binding protein [30], held together by RNA-protein and protein-protein interactions. Therefore, stress granules are considered to be repositories of mRNA translation machineries during stress. Some study suggested that stress granules regulate mRNA stability and translation by selectively recruiting mRNA to stress granules. For example, upon heat shock stress, the translation is reprogrammed towards a stress-specific state, including slowing down housekeeping protein translation and upregulating production of heat shock proteins such as molecular chaperones [65]. Consistent with that, the mRNAs encoding heat shock proteins are found to be excluded from the pool of mRNA recruited to stress granules [34]. Stress granules exhibit a dynamic feature that some component RNA-binding proteins regulating mRNA translation and decay can shuttle in and out stress granules [35]. These evidence points to an equilibrium between stress granules and translation machinery (polysomes) depending on the environmental condition. In this way, stress granules serve as a mRNA regulatory center. Nevertheless, there is no clear evidence showing that stress granule formation is adaptive.

1.0.4 Quinary structure

Before stress granule formation or under mild stress where stress granule does not form, many proteins already form condensed assemblies [66, 57]. Those assemblies fall into the classification of quinary structure [16], which refers to macromolecular assemblies of interacting biomolecules and therefore beyond the four lower levels of protein structures (primary-quaternary). The defining feature of quinary structure is a lack of fixed stoichiometry as the interactions are usually transient and weak, and quinary structures can undergo rapid kinetics of association and dissociation. These weak interactions are also termed quinary in-

teractions and are considered the driving force of protein condensation (when the interactions are multivalent), especially in the highly crowded cellular environment.

1.0.5 Poly(A)-binding protein structure and its condensation

An abundance of proteins associated with stress granules have been found to condense when purified *in vitro* [33, 51, 52]. Remarkably, Poly(A)-binding protein (PABP, Pab1 in budding yeast), an RNA-binding protein important for translational control [45] and a canonical stress granule marker that is consistently recruited to stress granules arising from different kinds of stresses [36, 2], has been found to condense under mild stress condition autonomously [57]. Upon heat shock or starvation-induced acidification, Pab1 condenses, demixing from the cytosol into foci, and is recruited to stress granules (Figure 1.4). Similarly, *in vitro*, purified Pab1 condenses above $T_{demix} = 39$ °C and at acidic pH independent of RNA.

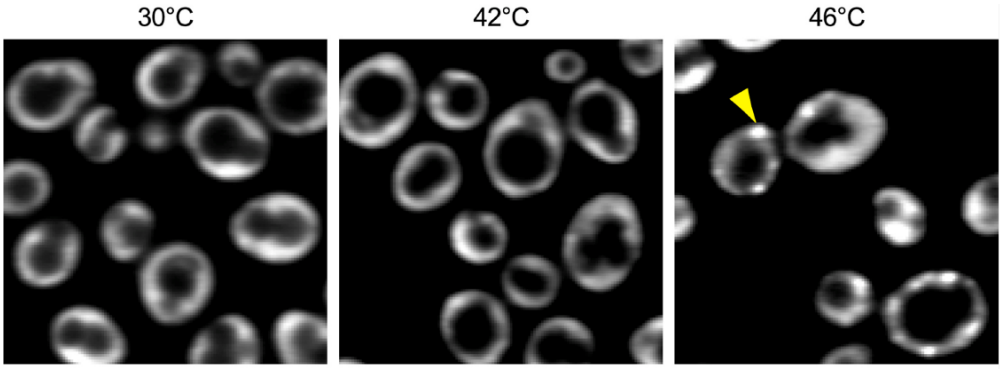


Figure 1.4: Pab1 condenses upon heat shock. Confocal fluorescent microscopy of yeast containing fluorescence labeled Pab1-mRuby2. The arrow indicates one of stress granules. Figure adapted from [57].

Pab1 contains four RNA recognition motifs (RRMs) connected by linkers, followed by an IDR termed P domain and a C-terminal peptide binding domain (CTD). Pab1's stress-triggered condensation has been shown to be an adaptive response of yeast that promotes cellular fitness during heat shock and energy-depletion stress [57]. Changes in the hydrophobicity of Pab1's 94 residue IDR P domain alters T_{demix} and growth levels. But condensation

is only compromised but not prevented by deletion of this IDR or the folded CTD [57], and the N-terminal half of Pab1 containing three of the four RNA Recognition Modules (RRM123) is sufficient for condensation (Figure 1.5) [57]. Pab1's condensation process results in the formation of irreversible gels and hence, is not a classic example of liquid phase separation. Cooling back to normal temperature or pH neutralization does not reverse Pab1's condensation. The condensates are not dissolved by high concentration of salt, though the condensation is an electrostatic-mediated process [57]. Nevertheless, *in vivo* Pab1 condensates are dispersed by the yeast disaggregation system (chaperones) directly and efficiently [69].

Dynamic Light Scattering (DLS) was used to monitor Pab1's condensation under a temperature ramp of 0.25 °C min^{-1}. (Figure 1.5B). A sharp increase in the hydrodynamic radius (R_h) denotes the onset of condensation, which is characterized by T_{demix}, the temperature where R_h is twice the pre-transition value (∼5 nm for Pab1, ∼3 nm for RRM123, the truncated, three RRM construct). Full length Pab1 and RRM123 demix at 39 and 46 °C, respectively (Figure 1.5B).

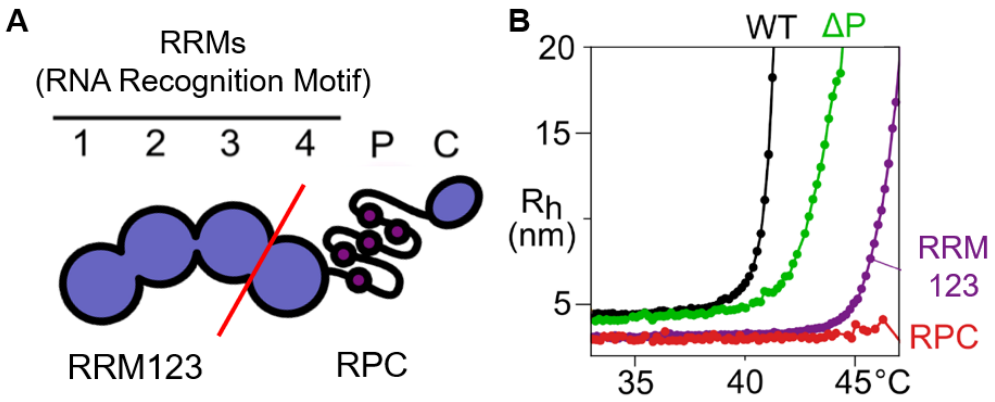


Figure 1.5: Pab1's structure and condensation monitored by DLS. (A) structure of Pab1 (B) DLS of Pab1, P domain deletion construct ΔP, and truncated three-domain constructs (RRM123 and RPC, structure shown in (A)). Figure adapted from [57].

Understanding how Pab1 senses thermal stress or acidification, transduces either stress,

and condenses would provide molecular details on how quinary assemblies are produced in response to stress, potentially facilitating the study of a wide range of systems. In addition, unlike most reported condensing protein system where IDR is required for condensation, Pab1's 3 RRMs (RRM123 construct) are sufficient to condense without the IDR P domain (Figure 1.5). It is unclear how folded RRMs instead of IDR mediate Pab1's condensation, and if it is a different mechanism from the canonical IDR-dependent model of condensation. Therefore, understanding the molecular mechanism of Pab1's condensation will complement the knowledge of both fields of stress biology and biomolecular condensates.

Molecular level details regarding how Pab1 condenses upon heating or pH drop remained unknown, partially due to that no readily available approach can probe the Pab1 condensates with high structural resolution. A lack of methodology to obtain such structural information is also an outstanding issue for the field, as it obstructs a deep understanding of the protein condensation mechanism. While mutational studies can identify some key structural elements for condensation, it is neither high throughput nor high resolution. The central challenge for conducting a high-resolution analysis of biomolecular condensates is that condensation itself impedes the acquisition of structural information by the usual set of biophysical approaches such as NMR. A biophysical method that can directly probe bimolecular condensates would benefit the study of bimolecular condensates from a structural perspective.

In this work, we aimed to investigate the molecular mechanism of the stress-triggered condensation of Pab1 by answering those questions:

1. How does Pab1 sense thermal stress or acidification through its structure and condense?

2. What structural changes does Pab1 undergo upon condensation?

3. How do folded RRMs mediate condensation as RRM123 is sufficient to condense independent of Pab1's IDR P domain? Do RRMs (and CTD) remain folded upon condensation?

4. Are pH- and Temperature-induced condensations of the same mechanism?

To answer above questions, we developed a protocol to apply Hydrogen-Deuterium Exchange/Mass Spectrometry (HDX-MS) to Pab1 condensates which provides H-bonding profile of the condensates and reveals structural events upon Pab1's condensation. In combination with other biophysical measurements that provide complementary information, our multi-facet study identifies the molecular factors encoded in Pab1 structure that enable it to transduce cellular stress signals and condense.

In Chapter 2, I present data on HDX-MS of both monomeric and condensates form of Pab1, which reveals that RRMs partially unfold upon condensation. In addition, RRMs have different levels and profiles of unfolding, where RRM3 remains largely folded. HDX data of condensates exhibits high level of heterogeneity that there are more than one population which have different HDX kinetics, which points to a structural diversity and interaction heterogeneity in the condensates. Another important observation is that pH- and Temperature-induced condensates are structurally similar.

In Chapter 3, we delineate that RRMs participate differently in condensation. Specifically, RRMs have different activation temperatures above which they partially unfold and participate in condensation. From this observation we proposed a "sequential activation" model for Pab1's condensation. We show that the activation is required for an RRM to be engaged in condensation by co-demixing experiments, a mechanism we term "thermodynamic specificity" wherein an activated RRM is more likely to interact with other activated RRMs rather than inactivated ones.

In Chapter 4, I discuss other molecular factors of Pab1's condensation, including histidines' titration, nucleation effect, intramolecular inhibition, and linkers' flexibility. Importantly, we found that histidines partly confer Pab1's sensitivity to pH. SAXS studies show that Pab1's linkers are flexible, and the condensation is not auto-inhibited. In correspondence, FRET confirms that Pab1 exhibits no significant large-scale conformational change upon condensation. Next, I present data from simulation for testing the mode of RRM interactions. The preliminary data suggests non-specific interactions between RRMs. In Chapter

5, I summarize the findings and conclusions described in previous chapters and address some open questions and future directions.

Overall, our multi-facet study identifies the molecular factors encoded in Pab1 structure that enable it to transduce cellular stress signals and condense. The sequential activation model and thermodynamic specificity also add to the knowledge of protein condensation as a novel mechanism. Finally, the successful application of HDX-MS to Pab1 condensates extends the territory of HDX-MS to the new field of biomolecular condensates, where HDX-MS will serve as a useful high-resolution analysis tool.

CHAPTER 2

APPLY HDX-MS TO PROBE PAB1 CONDENSATES

We applied HDX-MS to monomeric Pab1 and pH- and Temperature-induced Pab1 condensates. According to HDX-MS, 4 RRMs have different stabilities and RRM3 has the highest stability. In the condensates, extensive regions were destabilized and the individual RRMs exhibited different levels of partial unfolding with RRM3 remaining largely folded. HDX of condensates exhibited a high level of heterogeneity on most regions where two or more populations with distinct HDX kinetics were observed. Some regions became more stable than in the monomeric state. We also observed a similarity between pH- and Temperature-induced Pab1 condensates. Overall, regions throughout Pab1 exhibited changes in HDX behavior, mostly destabilizing, and considerable heterogeneity was observed. The HDX-MS results suggest a high degree of structural diversity and interaction heterogeneity in the condensates.

2.1 Introduction to HDX

In the recent decade, mass spectrometry has emerged as a useful analysis tool given its high sensitivity and versatility. Specially, it can be applied to complex biological samples including membrane protein, aggregates, and other challenging protein systems [37, 41, 50, 64], as the mass information is encoded in the protein itself and retained during downstream sample handling. Therefore, Hydrogen-Deuterium Exchange/Mass Spectrometry (HDX-MS) provides an option for biomolecular condensates as it can probe the hydrogen bonding network even in complex insoluble milieus and provide thermodynamic and structural information at the residue level [70, 20]. It can identify regions that are folded or unstructured directly, distinguishing freely exposed amides from amides protected by structures.

The principle of Hydrogen-Deuterium Exchange is that protein backbone amide proton can exchange with the proton from solvent due to the dynamic breakage and reformation of

the hydrogen bonding network. When the protein is subjected to D_2O, the proton will exchange with deuterium from solvent. From the Linderstrom-Lang model for HDX [20], when the protein backbone amide forms a protein-protein H-bond, it is exchange-incompetent. The H-bond can be broken upon a conformational change through thermal fluctuations and become available for exchange:

$$H_{closed} \underset{k_{close}}{\overset{k_{open}}{\rightleftharpoons}} H_{open} \xrightarrow{k_{chem}} D$$

Here, k_{chem} is the chemical rate for a free peptide, which is dependent on the peptide sequence and exchange conditions (pH, temperature) [4]. The steady-state solution is:

$$k_{obs} = \frac{k_{open} k_{chem}}{k_{open} + k_{close} + k_{chem}}$$

Depending on the relative rates between k_{close} and k_{chem}, HDX falls into two regimes yielding different information.

When $k_{close} \ll k_{chem}$, $k_{obs} = k_{open}$, which is termed EX1 regime where the observed rate is governed by the opening rate.

When $k_{close} \gg k_{chem}$,

$$k_{obs} = \frac{k_{chem} k_{open}}{k_{close} + k_{open}} = k_{chem} \frac{1}{K_{eq} + 1}$$

In this so-called EX2 regime, the observed rate is determined by the equilibrium constant between the open and closed states in combination with k_{chem}. When the amide proton is H-bonded, the k_{obs} will be slowed compared to k_{chem} by how much the structural stability is.

$$\text{Protection Factor} = \text{PF} = \frac{k_{chem}}{k_{obs}} = K_{eq} + 1$$

$$\Delta G = -RT \ lnK_{eq}$$

As k_{chem} can be reliably predicted for a given condition [4], from k_{obs} one will be able to tell the structural stability under EX2 regime. For example, a protection factor (PF) equal

to 1 when the region is exchanging at k_{chem}, indicates a fully unstructured region (Figure 2.1A). Another example is that the difference of HDX rates between two states reflects the difference of their stabilities. A 10 times difference of HDX rate corresponds to a $\Delta\Delta G$ of 1.36 kcal mol^{-1} (Figure 2.1B). Given the above capabilities, HDX is frequently applied for antibody epitope mapping and studying ligand binding effect and folding kinetics [31].

HDX follows first-order kinetics. The D uptake of each residue will be sigmoid-shaped on a log(t) scale, and the reaction constant can be calculated from the half time $t_{1/2}$. The D uptake of a peptide will be the sum of all D uptakes of each residue.

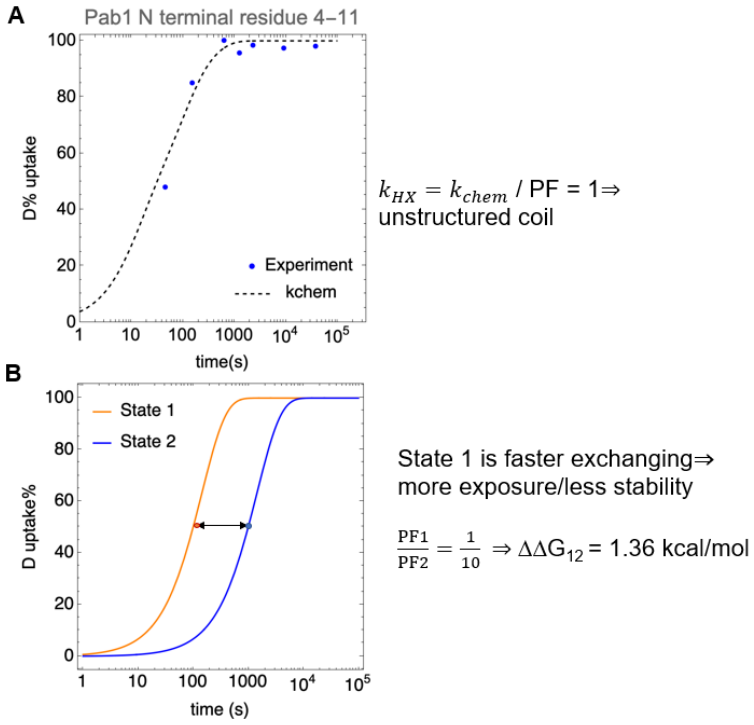


Figure 2.1: HDX provides thermodynamic information under EX2 regime. (A) A region exchanging at k_{chem} (PF = 1) is an unstructured coil. (B) Difference of HDX rate tells difference of structural stability quantitatively.

There are different ways to distinguish EX1 and EX2 regimes, including different pH dependence (k_{obs} under EX2 regime will change according to the change in k_{chem}, but not for EX1 regime, assuming K_{eq} and k_{open} are pH insensitive) and different isotopic distribution

behavior on mass spectra.

In the EX2 regime where multiple opening events occur before a given amide proton is likely to exchange, the mass envelope shifts continuously to higher molecular weight with an isotope distribution that satisfies the binomial distribution reflecting the stochasticity and independent exchange behavior of each amide protein (and other natural isotopes). For ideal EX1 exchange kinetics, however, there is the concerted exchange of multiple amides in an opening event that results in the decrease of one peak and a commensurate increase in another peak centered at a molecular weight that is higher by the number of exchanged sites. One can have mixed behavior with some sites exchanging with EX2 kinetics and others with EX1 kinetics. Generally, each peak maintains its amplitude while it continuously shifts to higher MW reflecting the EX2 kinetics while a more drastic shift in populations from one peak to the other can occur reflecting EX1 kinetics.

HDX is frequently coupled with NMR as a readout to study protein folding [21]. While HDX-NMR can theoretically provide residue-level information, there are limitations of this technique including: 1) NMR has an intrinsic restriction on the protein size (usually < 250 residues); 2) NMR requires a large amount of protein sample; 3) NMR only works with soluble sample (except solid-state NMR) and can not handle sample with more complex composition; 4) the peak assignment of residues can be difficult and requires a large amount of differently isotope-labeled sample.

Mass spectrometry addresses these issues of NMR based on its high sensitivity and consequently low sample amount needed, capability of protein sequence identification, and minimal restriction of protein size and sample composition. Therefore, the applications of HDX-MS to a range of different protein systems including membrane protein, protein aggregates and antibody in complex with ligand have a significant growth in the past decade (reviewed in [31]). In HDX-MS, high sequence coverage and resolution are achieved by a fragmentation approach [48], where many overlapping peptides are generated by proteolysis (done either by mixing with soluble protease or flowing through an in-line immobilized protease column

[48]), then separated by HPLC, and eventually read out on MS. By comparing the deuteration level of many overlapping peptides, one will be able to calculate the D uptake at a higher resolution, ideally at residue-level.

2.2 Results

2.2.1 HDX-MS methodology

To investigate the structural changes Pab1 undergo upon condensation, we conducted HDX-MS on three samples, monomeric (soluble) Pab1, pH-induced condensates (pH 4.50 for 30 minutes, 30 °C), and Temperature-induced (pH 6.80, 46 °C for 20 minutes) condensates (Figure 2.2). To initiate HDX labeling, the sample was diluted 29-fold into a D_2O solution (pD_{read} 6.00 on ice). The samples were allowed to exchange between 30 and 24,000 seconds followed by quenching at pH 2.5. 8 M urea was included in the quench solution to solubilize the condensates and any potential aggregates (Figure 2.3). Quenched samples were subjected to LC-MS analysis (Figure 2.4) with an in-line protease column (pepsin and a fungal protease) [48].

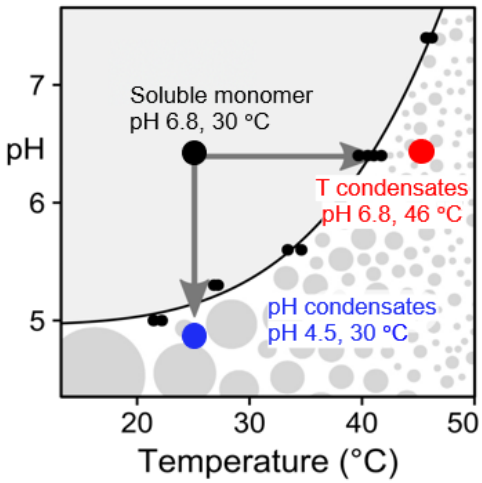


Figure 2.2: Preparation of three samples, including soluble/monomeric Pab1 and two different condensates, triggered either by pH or temperature.

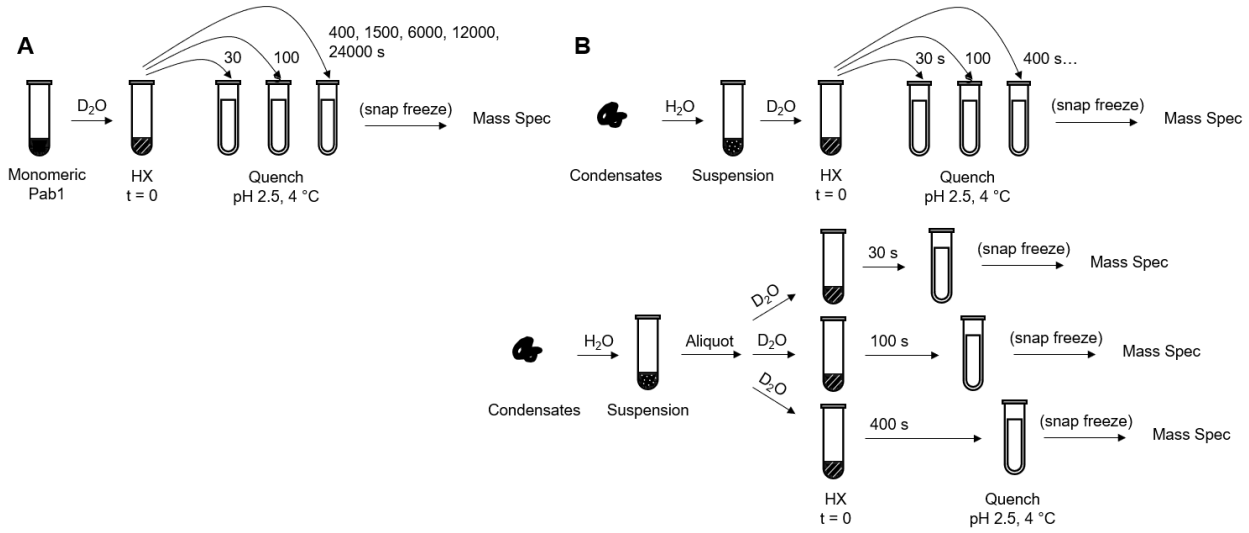


Figure 2.3: HDX labeling protocols. (A) Preparation of monomeric Pab1 HDX samples. (B) Preparation of Pab1 condensates HDX samples. The samples at each time point for each type of condensate are prepared either from a single bulk exchanging sample or from samples exchanging separately.

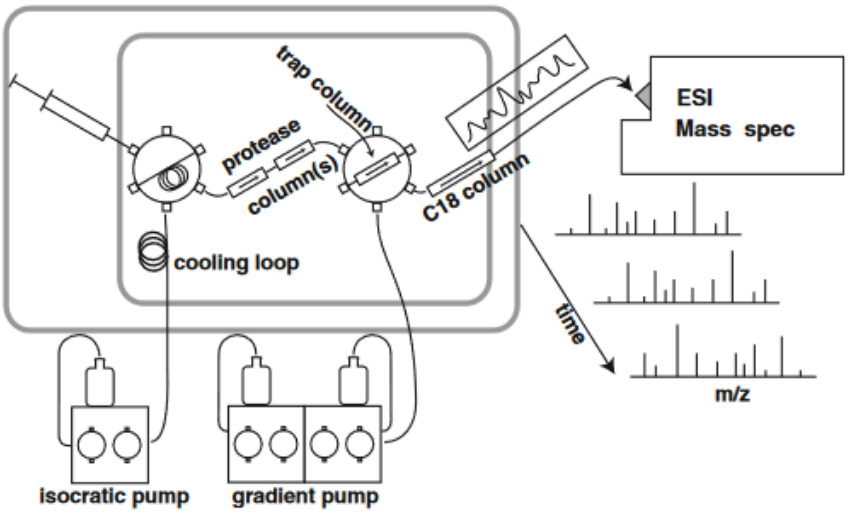


Figure 2.4: Hardware of the LC system for HDX-MS analysis. Quenched HDX sample is injected into the on-line flow system and flows through an immobilized acid protease column for proteolysis. Digested peptides are directed through a second valve to a C8 trap column. After sufficient flow time for desalting, a water/acetoniltrile gradient elutes peptides through a C18 analytical column for separation, then to the Mass Spec for readout. Figure is from [48].

2.2.2 HDX of monomeric Pab1 indicates RRMs have different stabilities

Our peptide map had 99% sequence coverage and an average redundancy of 7.1 (Figure 2.5 and 2.6). For monomeric Pab1, the HDX pattern aligned with the structural boundaries of the six domains. The five folded domains had peptides that exchanged at rates, k_{obs}, considerably slower than the intrinsic chemical exchange rate, k_{chem}, expected for an unstructured protein [20]. The regions known to be disordered, including the N- and C-termini, inter-domain linkers, and the proline rich P domain, exchanged at rates close to k_{chem}.

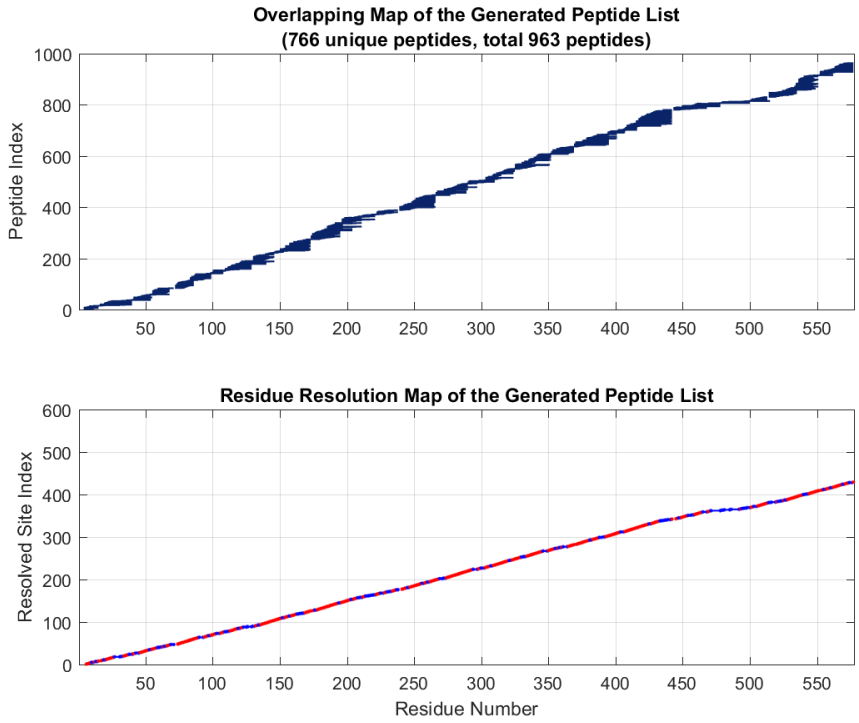


Figure 2.5: Pab1 MS/MS peptide coverage and resolution map. Red indicates a resolved site.

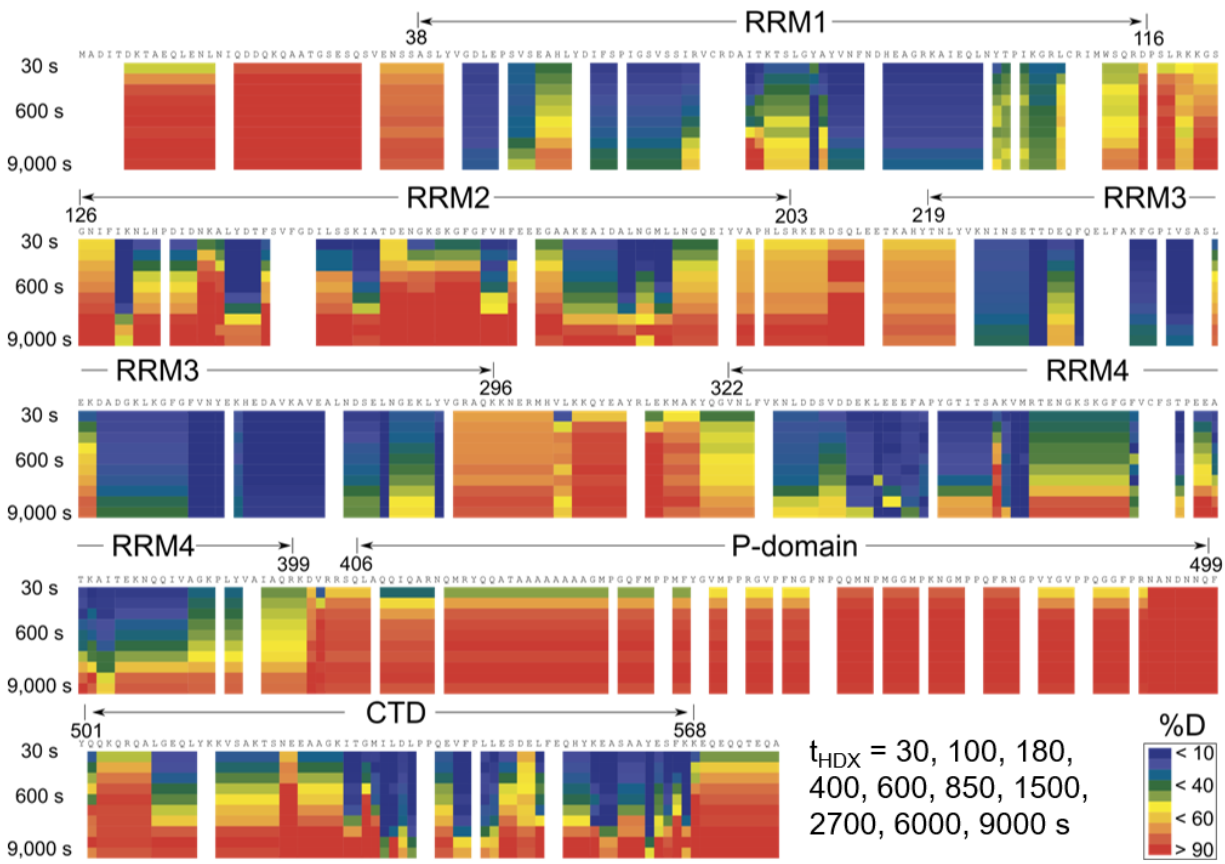


Figure 2.6: HDX heat map of monomeric (soluble) Pab1. Horizontal axis represents Pab1 sequence and vertical axis is increasing HDX timepoints at 30,100, ..., 9000 s. Color represents level of deuterium uptake. Boundaries of Pab1's six domains are marked on plot. Plot made with HDExaminer.

The four RRMs had different exchange behavior indicating that they have different stabilities. For RRM1, 2, and 4, most peptides exchanged a factor of 10^3-10^5 slower than k_{chem}. This ratio, termed protection factor (PF), corresponds to a global stability for each RRM of $RT\ lnPF$ = 4-7 kcal mol^{-1} (assuming HX is occurring in the thermodynamic EX2 limit [20], supplementary information). Many of the peptides from RRM3 exchanged even slower and have less than 50% deuteration even after 24,000 s (Figure 2.6 and 2.7). Accordingly, we estimate the stability of RRM3 to be greater than 7 kcal mol^{-1}.

As an example, the α2-helical region from 4 RRMs had different D uptake kinetics, where RRM3 exchanged at least 10 times slower than the other 3 RRMs (Figure 2.7).

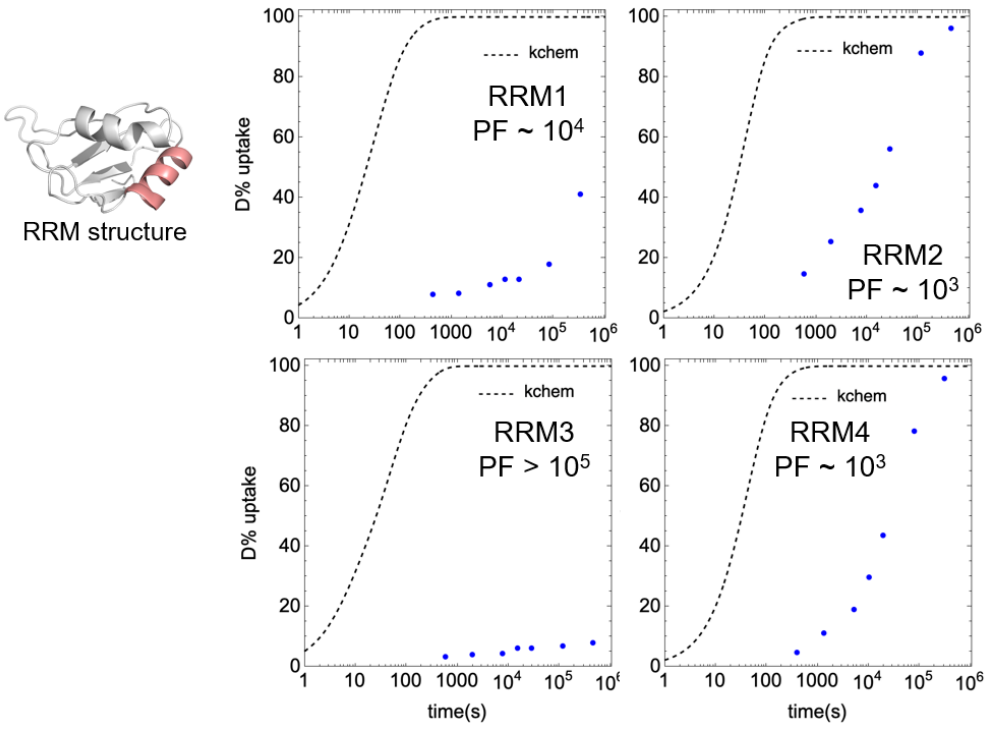


Figure 2.7: Deuterium uptake curves of peptides from α2-helices of 4 RRMs. RRM1 structure is shown with α2-helix highlighted. Estimated PFs are indicated on plots. D uptake level was normalized to an all-D control to account for back exchange.

2.2.3 HDX of Pab1 condensates exhibits heterogeneity and partial unfolding

In the monomeric state, each peptide exchanged as a single population. However, the HDX behavior of the two types of condensates exhibited significant yet similar heterogeneity, which persisted across bio-replicates (Figure 2.10 and 2.12). Many peptides had obvious bimodal mass spectrum envelopes with one subpopulation exchanging at or close to k_{chem}, potentially being unfolded, as well as another subpopulation exchanging at comparable or even slower HDX rates than observed in the monomeric state (Figure 2.8, middle&right panel). This bimodal behavior can be explained by the partial unfolding of regions of the protein in a subpopulation of the molecules. For a few peptides, however, both populations were slowly

exchanging, and the slower population exchanged even slower than observed for soluble Pab1. We attribute this heightened slowing to new interactions between folded Pab1 and other molecules in the condensate (Figure 2.8, bottom panel). Because of the heterogeneity, we analyzed the data using the bimodal fitting option available in the HDExaminer software to extract the time dependent $\%D$ buildup curves of each population and their relative fractions over time.

Multiple regions in RRM1, 2, and 4 became destabilized in the condensates (Figure 2.10 and 2.11). The regions included the helices, hairpins and β-strands, depending on the RRM. In addition, the RRMs displayed variable levels of heterogeneity (Figure 2.12). The fast-exchanging population fraction ranged from close to zero for the most stable parts of RRM3 to greater than 70% for the α2-helical region in RRM4 (Figure 2.12).

Different RRMs have different unfolding profiles. Most of the destabilization observed in RRM1 and RRM4 occurred in the α2-helical regions (60% for RRM1, 70% for RRM4) and L1 hairpin (60% for RRM1, 50% for RRM4). The moderate level of unfolding observed in RRM2 occurred at the L2 hairpin instead of the helices. In contrast, RRM1's L2 hairpin became more protected compared to the monomer state. Moreover, L3 hairpins of RRM1 and RRM4 could also become unfolded but RRM2's L3 showed little difference from the monomeric rate (Figure 2.11 and 5.2).

Compared to the other RRMs, the D uptake behavior of RRM3 was closer to that of the Pab1 monomer and the fast exchanging population was reduced to under 30% for all RRM3 peptides, implying that condensation had less of an effect on this domain. In RRM1, 2 and 4, some β-strands also unfolded upon condensation (Figure 2.10 and 2.12). The β1 and β3 strands along with the L3 loop of RRM are known to interact with single strand poly(A) RNA bound to RRM1-2 [46]. The loss of these structures in the condensates may explain the previous finding that RNA is ejected from Pab1 upon condensation [57].

Some stable regions exchanged $\sim$100-fold slower than their monomeric counterparts (e.g., Peptide_{73-82}, Figure 2.8 bottom panel). This stabilization in the condensates is likely due to

new interactions with either other folded domains or regions that undergo partial unfolding. Notably, the linkers between RRM2 & RRM3 and RRM3 & RRM4 were unstructured in monomeric Pab1 but became at least 100-fold more protected in the condensates, potentially H-bonding to their counterparts or to other unfolded regions on different molecules. Based on these results, we conclude that Pab1 molecules can adopt multiple, partially unfolded conformations and form a variety of inter-chain interactions in the condensates.

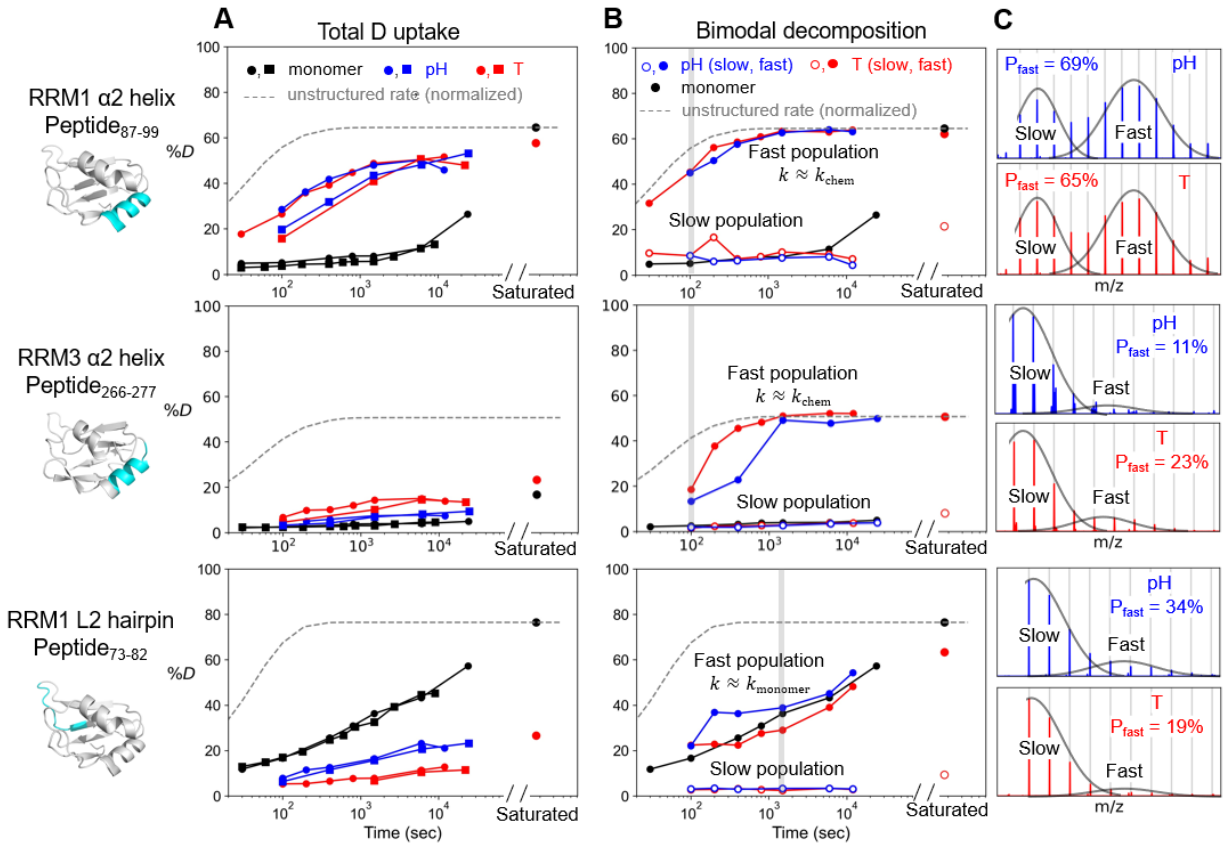


Figure 2.8: Uptake curves of RRM1 and RRM3 α2-helices, and RRM1 L2 hairpin. (A) Total D uptake curve for the combined fast and slowly exchanging populations (data from 2 replicates are shown). (B) Individual D uptake curves for the fast and the slow populations obtained using a bimodal fitting procedure. For both plots, unstructured rate (k_{chem}) is normalized to back exchange level. (C) Bimodal mass spectrum envelopes indicating two (or more) different populations taken from timepoint 400, 400, 1500 s, respectively. Corresponding plots for RRM2 and RRM4 α2-helices are presented in Figure 2.9.

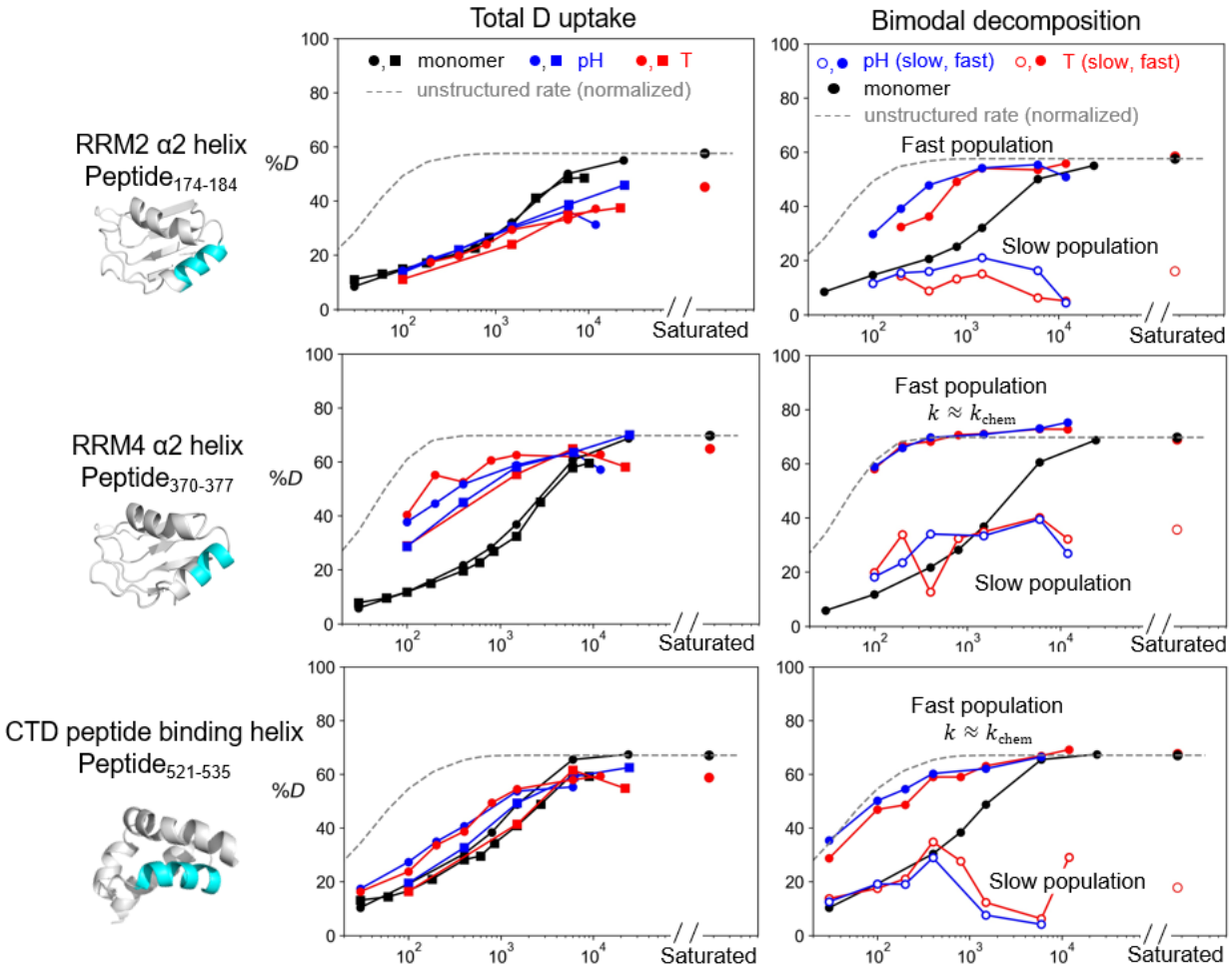


Figure 2.9: Uptake curves of RRM2, RRM4 α-2 helices, and CTD peptide binding helix. Left is total D uptake curve for the combined fast and slowly exchanging populations. Right is D uptake curves for the fast and the slow populations obtained using a bimodal fitting procedure.

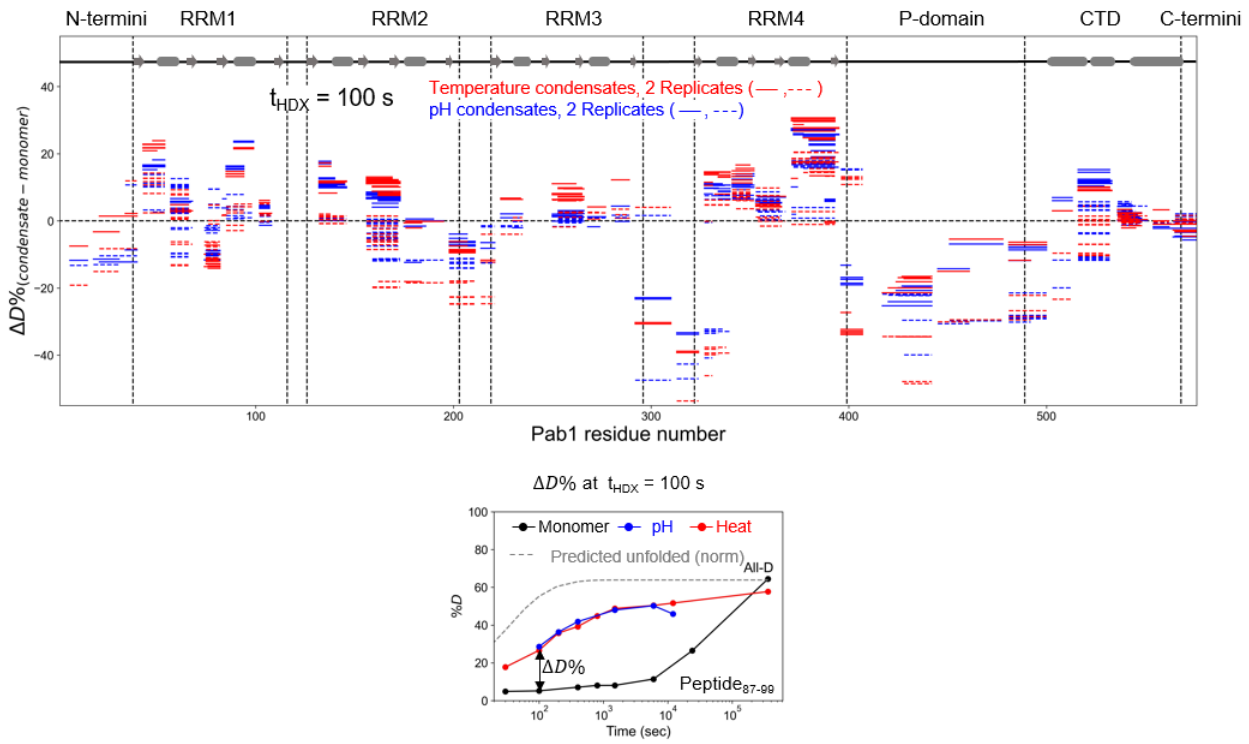


Figure 2.10: Wood plot for HDX of Temperature- and pH-induced condensates compared to monomeric Pab1. $\%D$ differences of both condensates compared to monomeric state at $t_{HDX} = 100$ s are plottedfor each peptide.

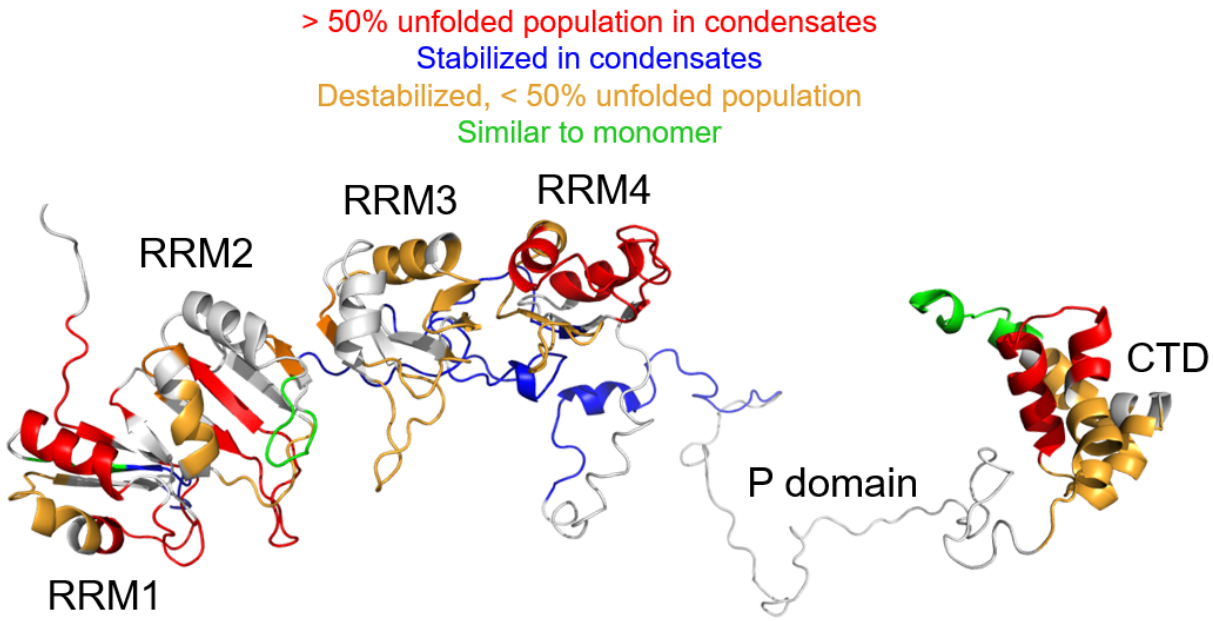


Figure 2.11: HDX mapped onto Pab1 structure.

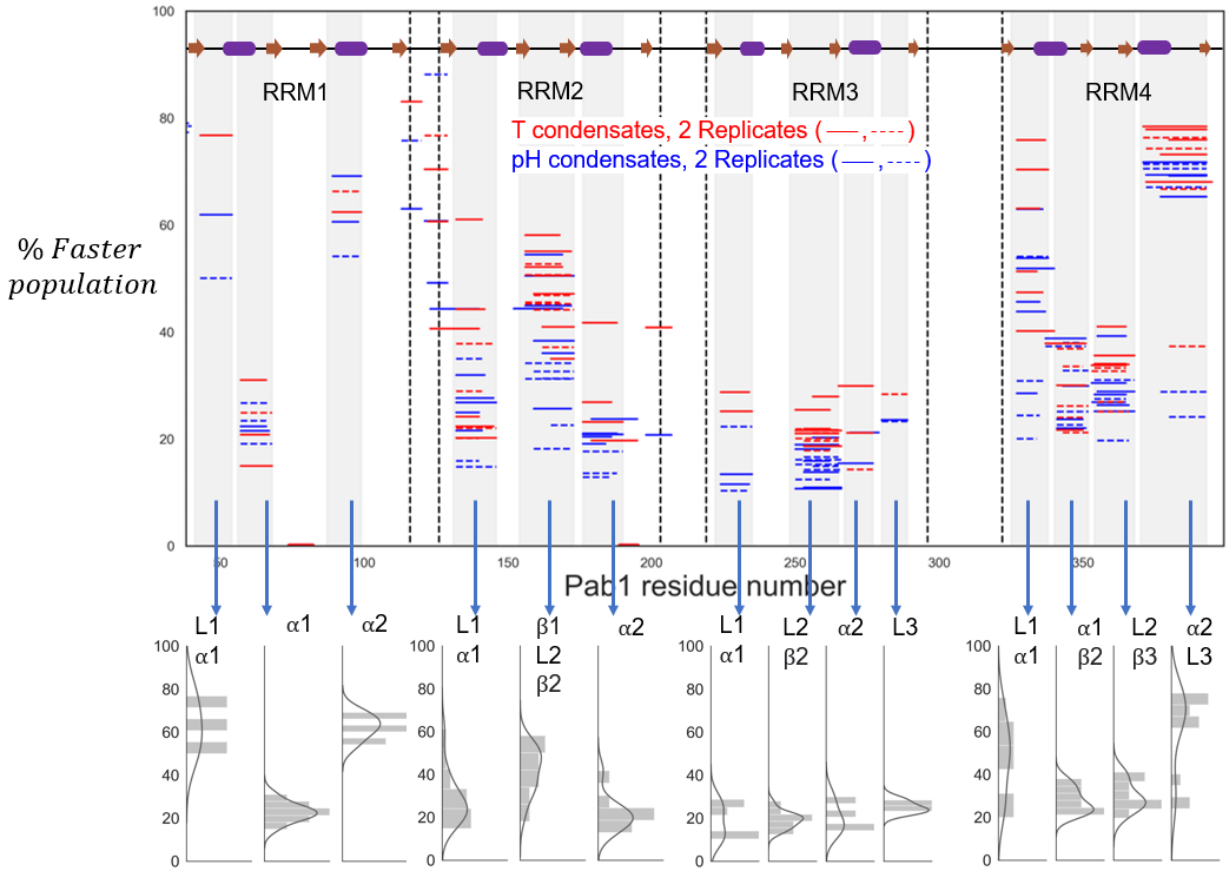


Figure 2.12: Heterogeneity map of Temperature- and pH-induced condensates. Fraction of the fast exchanging population for regions of RRMs that become unfolded in the condensates are plotted for each peptide. Peptides from linker region are not shown as they are unfolded in the monomeric state. Secondary structures are depicted on top. Density distributions for secondary structure elements are shown in lower panel.

We observed less change in HDX upon condensation for the P domain and C-terminal domain (CTD), although heterogeneity still can be found in some regions (Figure 2.9, bottom panel). HDX was slowed by 5-fold for ∼50% of the P domain molecules across residues 420-440 (Figure 2.10), spanning the region of eight contiguous alanine residues, implying that it formed interactions in the condensates. Potentially, this region became helical in the condensates. The HDX behavior of CTD's helix 2 ($\text{Peptide}_{521-535}$, part of the CTD's peptide binding site [39]) indicates that it was unfolded in the majority of the molecules (Figure 2.9 bottom panel). Some other regions of the CTD had slightly accelerated HDX

(Figure 2.10), potentially resulting from the disruption of the hydrophobic core due to the loss of the helix 2. Together with the observation that RRM123 can condense on its own whereas the RPC (RRM4-P domain-CTD) construct does not, we conclude that P domain and CTD participate in the condensation process to some degree even though they are not essential.

2.2.4 Isolated RRMs are well folded above T_{demix}

To investigate whether the disorder observed in the condensates is result of the intrinsic thermal instability of the individual domains, we collected ^{1}H-^{15}N HSQC spectra for the individual RRM1, 2, and 3 domains at 35 °C and 45 °C which are below and above Pab1's T_{demix} of 39 °C respectively, at pH 6.8. At both temperatures, the NMR spectra for the 3 individual RRM domains remained well-dispersed, the hallmark of a folded protein, and were nearly unchanged with no evidence of any conformation change upon heating (Figure 2.13). Furthermore, the peak volumes did not measurably change implying that all the protein molecules stayed in solution, determined by a comparison between the ^{1}H integral of peaks of methyl groups on residue side chains and reference compound (tetramethylsilane, TMS). The NMR T_1/T_2 measurements of RRM1, which are sensitive to tumbling rates [9], were consistent at both temperatures. At 35 °C, $R_1 = 1.85 \pm 0.076\ s^{-1}$, $R_2 = 5.88 \pm 0.61\ s^{-1}$, $t_c = 5.7 \pm 0.25\ ns$. At 45 °C, $R_1 = 2.06 \pm 0.017\ s^{-1}$, $R_2 = 4.54 \pm 1.54\ s^{-1}$, $t_c = 5.1 \pm 0.56\ ns$. Those results yield an estimated molecular weight of ∼10 kDa, which indicates that RRM1, which is the most activable RRM, remains monomeric at both temperatures. Also, no new peaks appeared at 45 °C associated with rapid HDX and disorder in a water saturation-transfer measurement (CLEANEX-PM [29]). Finally, at pH 5.2 (0.2 units below pH_{demix} at 30 °C), RRM1's spectrum likewise remained similar to the pH 6.8 spectrum (Figure 2.13). In summary, the 3 individual RRM domains remain stably folded under conditions where Pab1 condenses.

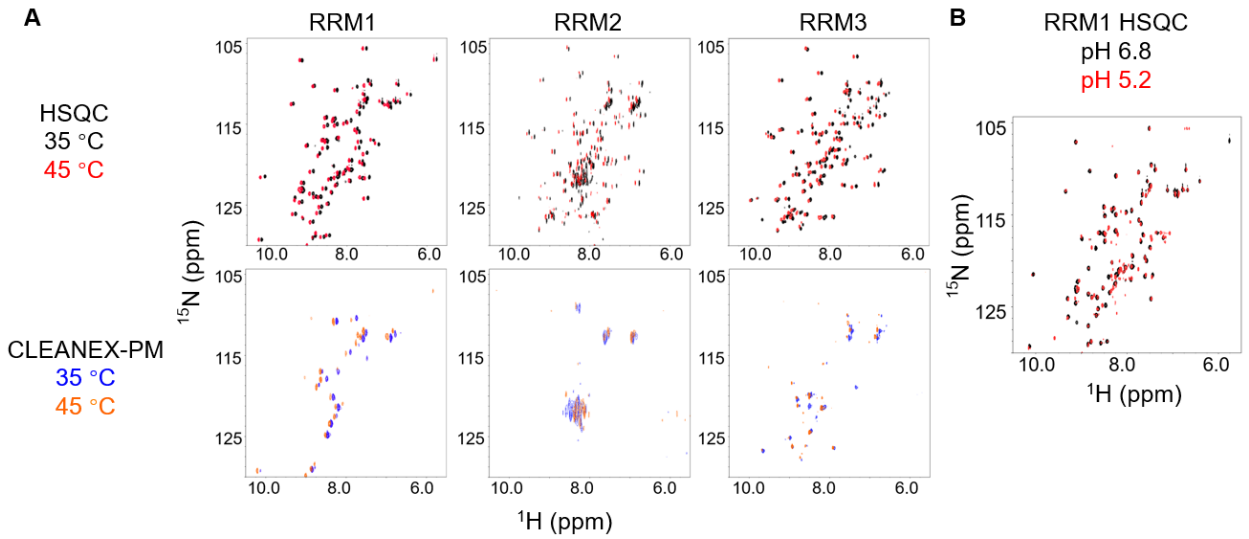


Figure 2.13: NMR shows that individual RRMs remain stable above Pab1's T_{demix}. (A) Overlaid ^{1}H-^{15}N NMR HSQC spectra (upper) of RRM1, 2 and 3, at 35 and 45 °C, pH 6.8, and corresponding CLEANEX spectra (lower) at the same conditions. (B) Overlaid ^{1}H-^{15}N NMR HSQC spectrum of RRM1 at pH 6.8 and pH 5.2, 30 °C.

2.2.5 Similarity between pH- and Temperature-induced condensates

We found great similarities between the pH- and Temperature-induced condensates (Figure 2.14, 2.10, 2.12). Their HDX patterns resembled each other. For most regions with heterogeneity, we found the two condensates were composed of the same set of underlying populations. This finding indicates that Pab1 undergoes similar structural changes upon pH- or temperature-triggered condensation and the mechanisms are likely the same. Nonetheless, some regions can respond to pH and temperature with different sensitivity. RRM3 is more responsive to temperature compared to pH, as Temperature-induced condensates have more of the fast population (30%) compared to pH-induced condensates (below 15%) (Figure 2.12). The similarity between the two condensates suggests that the mechanisms of pH- and Temperature-induced condensation are likely the same, which enables Pab1 to respond to both stresses. We propose that pH triggers Pab1's condensation partly through histidine titration, where the destabilizing Coulombic repulsion may cause unfolding of the region and subsequent condensation (discussed in Chapter 4).

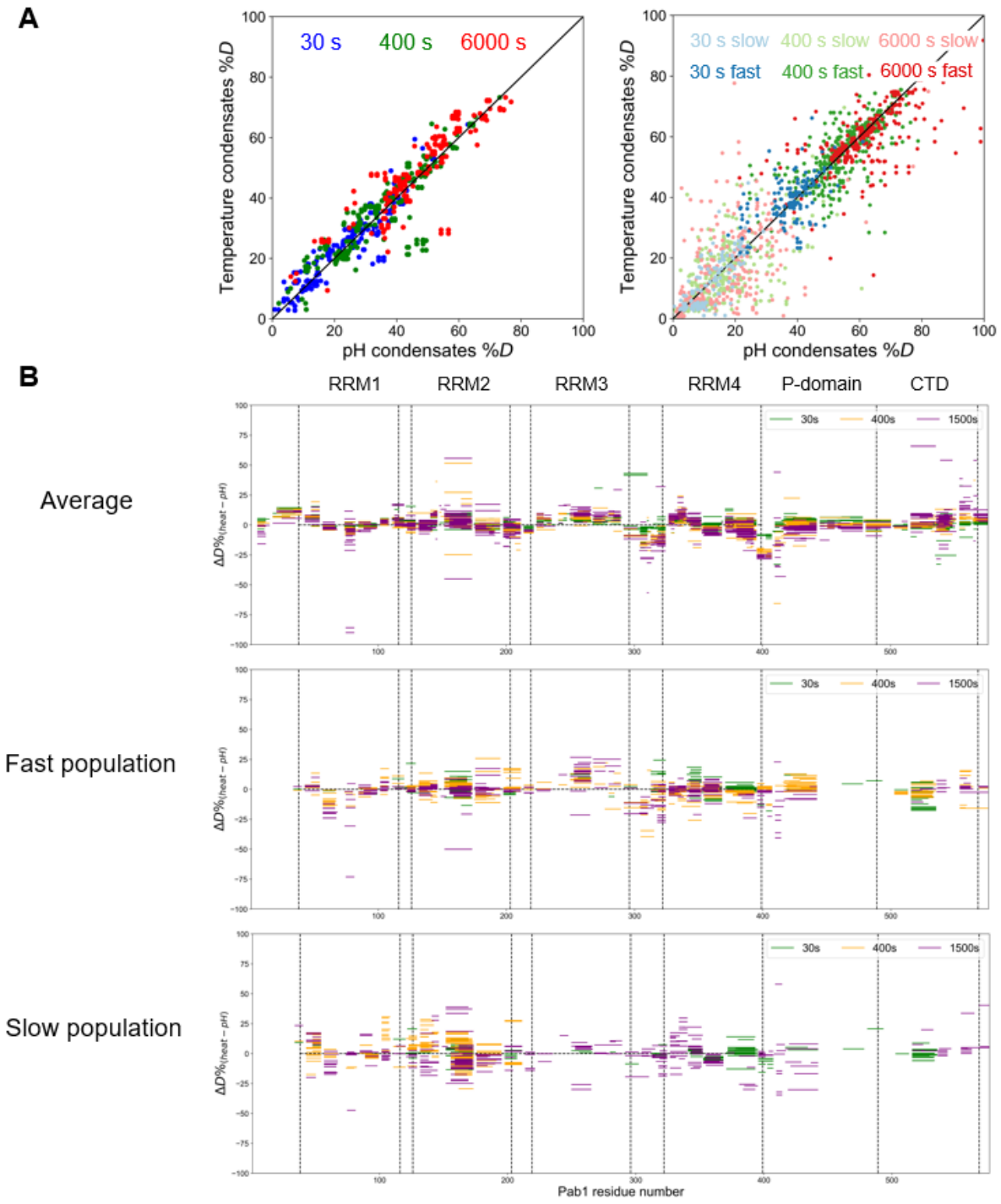


Figure 2.14: HDX of pH- and Temperature-induced condensates are similar. (A) Scatter plot comparing pH and Temperature condensates at 3 exchange times. Left is $\%D$ difference of total uptake (sum of the fast and slow exchanging populations for regions with heterogeneity). Right panel is $\%D$ difference of fast or slowly exchanging population. (B) $\%D$ difference between Temperature- and pH-induced condensates at 3 exchange times. $\%D$ difference of total uptake, fast exchanging population, and slowly exchanging population individually are shown in three panels.

The similarity between the two condensates argues that Pab1 arrives at the same condensed phase regardless of by means of temperature increase or pH drop, which enables cells to perform stress response to different types of stresses, highlighting Pab1's role as a key cellular stress sensor.

2.2.6 Data reproducibility

We prepared 2 bio-replicates where the proteins were from cells grown separately. The scatter plot together with the wood plot and heterogeneity map, where data from 2 replicates are plotted together indicate that the data is highly reproducible (Figure 2.15). For some regions, two replicates have different heterogeneity where the $\%faster\ population$ of two replicates are different (Figure 2.12). Besides experimental error, we speculate that this difference can result from the quinary and non-stoichiometric feature of the condensation such that one region is not strictly required to unfold and mediate the condensation at the same level in each replicate.

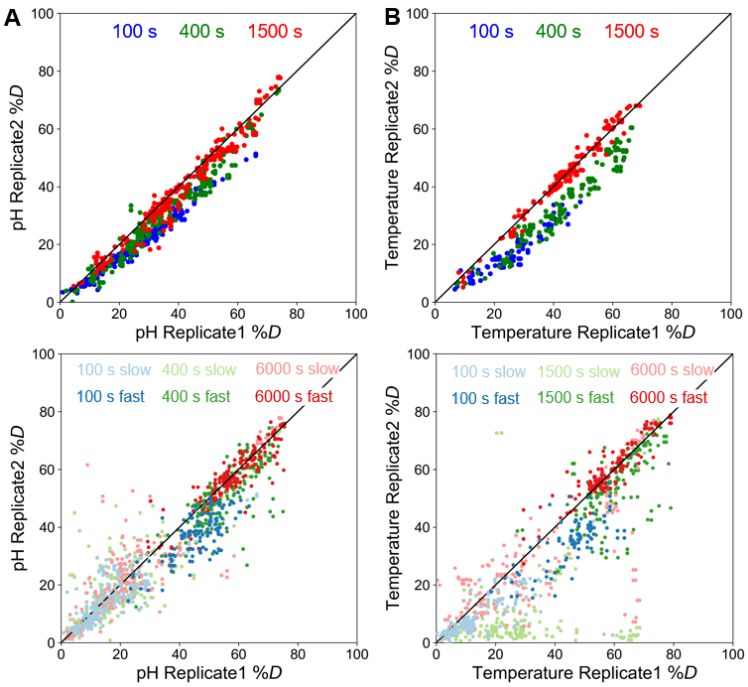


Figure 2.15: Scatter plot comparing 2 replicates of either pH- or Temperature-induced condensates at 3 exchange times. Upper is $\%D$ difference of total uptake (sum of the fast and slow exchanging populations for regions with heterogeneity). Bottom is $\%D$ difference of fast and slowly exchanging population individually.

2.3 Discussion

We conducted HDX-MS to probe hydrogen bonding in Pab1 condensates. It was a puzzle why Pab1 does not require its IDR, P domain, for condensation, but requires the structured RRMs. For condensing protein composed of both folded domain and IDR, the IDR is usually required for their condensation [17, 33, 51, 55, 24]. It was unclear how RRMs mediate Pab1's condensation and if the mechanism is distinct from LCR/IDR-mediated condensation. HDX-MS reveals partial unfolding of RRMs upon condensation, which illuminates a novel mechanism of how structured domains mediate condensation by partially unfolding.

We observed population heterogeneity on most regions in Pab1 condensates. This heterogeneity implies molecules are adopting different conformations in the condensates, including populations that are unfolded, similarly protected to, or more protected than the monomer.

Regions that show heterogeneity in the condensates, can have different component populations. The mostly observed populations are: 1) an unfolded population where $k_{obs} \sim k_{chem}$ 2) an unchanged population where $k_{obs} \sim k_{monomer}$ 3) a more protected population where $k_{obs} < k_{monomer}$. These underlying populations are suggesting that Pab1 molecules in the condensates adopt different structures with differing H-bond levels in certain regions. The molecules can retain the same H-bonding pattern as in the monomer, or break preexisting H-bonds to unfold, or form new intermolecular interactions, which is manifested in the heterogeneity observed in the HDX of the condensates.

Our result also highlights the potential of HDX-MS to study biomolecular condensates with high resolution in both space and energy. In Pab1's case, RRMs' unfolding only happens in the condensates. Therefore, the unfolding of RRMs upon condensation is not detectable by CD (on diluted Pab1) or NMR (on individual RRM), as those techniques cannot be directly applied to large proteins/condensates. HDX-MS can directly probe Pab1 condensates under pertinent conditions and provide high-resolution structural information across the whole sequence. Moreover, mass spectrometry clearly distinguishes different populations in the condensates, which will be a huge advantage to study other biomolecular condensates, as the heterogeneity may be a common feature of condensates systems, arising from the fact that condensation is often driven by quinary interactions which lack a fixed stoichiometry.

2.4 Supplementary information and methods

2.4.1 EX2 regime determination and discussion on heterogeneity

In the EX2 regime where multiple opening events occur before a given amide proton is likely to exchange, the mass envelope shifts continuously to higher molecular weight with an isotope distribution that satisfies the binomial distribution reflecting the stochasticity and independent exchange behavior of each amide protein (and other natural isotopes). As discussed in the introduction session, each peak maintains its amplitude while it continuously

shifts to higher molecular weight reflecting the EX2 kinetics while a more drastic shift in populations from one peak to the other can occur reflecting EX1 kinetics.

EX2 behavior was observed in our data, where the isotopic envelope adopted a binomial distribution that is convoluted with the distribution of deuterons on each amide and the distribution of other natural isotopes and moved to higher MW. We observed two distinct mass envelopes (or more for some peptides) in our HDX data for the condensates. This bimodal HDX data can be indicative of two structurally distinct populations that do not interconvert on the labeling timescale or from EX1 exchange kinetics with concerted exchange that results in the peak at lower MW decreasing in height while the peak at higher MW increasing by the same amount. Since peak heights remain relatively constant within each bio-replicate across different timepoints, we conclude that the two populations are largely due to static structural heterogeneity in Pab1 condensates.

2.4.2 Methods

Condensates preparation

Condensates were prepared by subjecting a 60 μM Pab1 stock to the condensing condition (pH 6.80, 46 °C for 20 minutes for T-condensates; pH 4.50, 30 °C for 30 minutes for pH-condensates). After treatment, the samples were centrifuged at 15,800 g for 10 min followed by 2x wash by buffer under the same centrifugation condition.

HDX-labeling

HDX was initiated by diluting samples 29-fold into D_2O buffer (50 mM sodium phosphate, 100 mM NaCl, pD_{read} 6.00) on ice. At corresponding HDX timepoint, quench buffer containing 600 mM Glycine, pH 2.5 and 8 M urea was added to quench HDX. Quenched sample was further diluted with buffer containing 600 mM Glycine, pH 2.5 to arrive at a final urea concentration of 2.67 M.

For condensate Replicate 1, condensates were resuspended and aliquoted and D_2O buffer was added to each condensate aliquot individually. At corresponding HDX timepoint, quench buffer containing 600 mM Glycine, pH 2.5 and 8 M urea was added to quench HDX and dissolve condensates. Quenched sample was further diluted with buffer containing 600 mM Glycine, pH 2.5 to arrive at a final urea concentration of 2.67 M.

For Replicate 2, condensates were resuspended and diluted into D_2O buffer as a bulk HDX reaction, then quenched at corresponding HDX timepoint using the same protocol as for the monomeric Pab1.

LC-MS

All samples were filtered by spinning for 30 seconds at 15,800 g at 4 °C in an Eppendorf 5415 R centrifuge fitted with a FA-45-24-11 rotor and injected into the loading loop of LC system. Sample was subjected to digestion by an in-line immobilized protease column of a mixture of Protease from Aspergillus Saitoi (Sigma-Aldrich P2143-5G) and porcine pepsin (Sigma-Aldrich P6887), at a flow rate of 100 μL/minute of buffer A (0.1% formic acid, pH brought to 2.5 by TFA). Protease column was prepared by immobilizing protease to resin (Thermo Scientific POROS 20 Al aldehyde activated resin 1602906) and packing into a column (IDEX Analytical Guard Column, C-130B). Digested peptides were trapped on a C8 column (TARGA C8 5 μm 1.0x5.0 mm Piccolo Column, TP-M501-C085) and desalted for 3 minutes at 100 μL/min. After desalting, the trap column was diverted to be in line with a Dionex Ultimate-3000 gradient pump (Thermo Fisher Scientific) flowing at 20 μL/min. A C18 analytical column (TARGA C18 0.5x50mm 120A, TS-05M5-C183) was in line to separate peptides upstream of Thermo Q Exactive mass spectrometer attached with the HESI-II (Thermo Fisher Scientific, cat# IQLAAEGABBFACNMAGY) probe. A 14-minute gradient (buffer A: 0.1% formic acid, buffer B: 0.1% formic acid in acetonitrile) from 10% to 60% buffer B was immediately started. MS data were collected from 2 to 13 minutes of the gradient. After the gradient, the pump was set to 90% B for 2 minutes, then down

to 10% B for a minute, followed by two trapezoid washes up to 90% B. Blank runs were inserted between sample runs where a mock sample with no protein was injected. Protease column was 2xwashed after each run by manually injecting a wash solution containing 1.5 M GdmCl, 4% Methanol and 0.8% Formic Acid.

Electrospray ionization was performed at 180 °C with the following parameters for Mass Spectrometry: spray voltage set to 3.2 kV, 1 microscan per scan, resolution 140,000, AGC target 3e6, minimum IT 100 ms, scan range 400-2000 m/z, dynamic exclusion 10 ppm, sheath gas flow rate 5, aux and sweep gas flow rates 0. For assignment MS/MS runs, data were acquired with the following parameters: data collected from 2 to 13 minutes of an identical gradient to MS runs detailed above, full MS as detailed above, dd-MS2: resolution: 17,500, automatic gain correction target: 1e5, maximum integration time: 200 ms, loop count: 10, MSX count: 1, TopN: 10, isolation window: 1.9, m/z isolation offset: 0.00, scan range: 200 to 2000 m/z, fixed first mass: 100.0 m/z, NCE / stepped: NCE 27, spectrum data type: profile.

HDX-MS data analysis

For assignment, MS/MS data was searched against a sequence database containing sequences of Pab1, protease, and other proteins that have been injected onto the LC-MS) using SearchGUI software (CompOmics Group). Search settings: unspecific cleavage, precursor charge 1-8, isotopes 0-1, precursor m/z tolerance 10.0 ppm, fragment m/z tolerance 0.5 Da, no post-translational modifications, peptide length 4-30. Searched result was imported into PeptideShaker (CompOmics Group) and further processed by EXMS2 (http://hx2.med.upenn.edu/EXms/) to generate a peptide list.

For HDX analysis, MS data together with peptide list was imported into HDExaminer 3.1 (Sierra Analytics) to fit peptide isotope distributions. For condensates where some peptides have distinct bimodal mass spectrum envelope, bimodal fit was used to determine the mass centroids of two individual populations. Downstream analysis and plotting were performed

in Jupyter Notebook. HDX data exported from HDExaminer was filtered by confidence score >= 0.88. For bimodality/heterogeneity analysis, a filter of peptide length>= 9 was applied that isotopic envelopes of both populations are separated enough for analysis.

Protein expression and purification

Protein expression and purification protocols were adapted with modification from [57]. BL21 *E. coli* cells transformed with an expression plasmid for N-terminally 8xHis-tagged protein constructs were grown in LB at 37 °C until the optical density at 600 nm (OD600) reached between 0.6 and 0.7. The flask was cooled down at room temperature for 30 minutes before being transferred to a 30 °C incubator. IPTG was added at a final concentration of 0.2 mM to induce protein expression. Bi-Cys mutant protein was expressed in Shuffle T7 Competent E. coli cells (NEB, cat# C3026J) and cells were grown at 30 °C instead of 37 °C.

Cells were harvested after four hours of induction and lysed by sonication on ice in buffer containing 20 mM HEPES at pH 6.7, 150 mM KCl, 5 mM imidazole, 1 mM PMSF and 0.1% Triton. Lysate was clarified at 13,000 g for 30 minutes and loaded onto a 5 mL HiTrap chelating HP column (Cytiva 17-0409) on a Bio-Rad FPLC system. Bound protein was eluted with an imidazole gradient. Fractions containing the target protein were pooled with b-mercaptoethanol and TEV protease and dialyzed into 20 mM HEPES pH 6.7, 150 mM KCl, 10% glycerol, to remove N-terminal TEV tags. TEV-cut fractions were pooled and loaded onto 2x1 mL HiTrap heparin HP column (Cytiva 17040601) for removal of nucleic acid contaminants with elution over a KCl gradient. Protein concentration was measured by absorbance at 280 nm.

NMR

NMR data was acquired either on Bruker AVANCE IIIHD 600 or a Bruker AVANCE III 500 NMR spectrometer equipped with a room temperature TXI probe with Topspin 3. Samples were exchanged to buffer at indicated pH (20 mM HEPES at pH 6.8, 150 mM KCl,

2.5 mM $MgCl_2$, 2 mM TCEP or 50 mM sodium acetate at pH 5.2, 100 mM KCl, 2 mM TCEP) and concentrated to > 150 μM. Mixing time for CLEANEX-PM measurements was 200 millisecond. Data was inspected in Topspin, processed with NMRFx and plotted with NMRviewJ.

CHAPTER 3

SEQUENTIAL ACTIVATION CONTROLS PAB1'S CONDENSATION

Based on the HDX-MS results that RRMs have different stabilities and exhibit different levels of unfolding in Pab1 condensates, we engineered triplet-RRM constructs which contain three copies of the same RRM. DLS on those constructs reveals that RRMs contribute differently to condensation, which complements HDX-MS results. An engineered disulfide-bonded construct indicates unfolding of RRM1's helices is an important early step of condensation. Those results point to that each class of RRMs becomes activated and capable of participating in the condensation process at different temperature thresholds. Co-dexmixing experiments show that the activation is a necessary step for an RRM to participate in condensation. Based on these results, we propose a "sequential activation" model where each RRM is activated at different temperatures, executes partial unfolding, and associates strongly only with other activated RRMs to form the condensate via a mechanism we term "thermodynamic specificity".

3.1 Results

3.1.1 Triplet RRMs indicate the RRMs types act differently

We investigated the extent to which each RRM promotes condensation by measuring the T_{demix} of triplet RRM constructs, termed RRM111, RRM222 and RRM333, which retain the native linkers (Figure 3.1). RRM111 precipitated even at room temperature whereas RRM123 and RRM222 demixed at 46 and 48 °C, respectively. RRM333 remained monomeric with an unchanging R_h even up to 50 °C. The relative order of T_{demix}, RRM111 $\ll$ RRM123 $<$ RRM222 $\ll$ RRM333, tracks with the relative stability of the individual RRMs identified by HDX. These results suggest that each of the RRMs becomes activated and capable of

participating in the condensation process at a different temperature.

This result distinguishes Pab1's condensation from most other condensation system with a lack of specificity. Here, different RRMs are encoded with different information for condensation in their sequences. Moreover, as swapping RRMs significantly alters T_{demix} while the linkers are kept the same, and RRM333 is non-condensing even it has the same linkers, we conclude that RRMs are the key to condensation rather than the linkers.

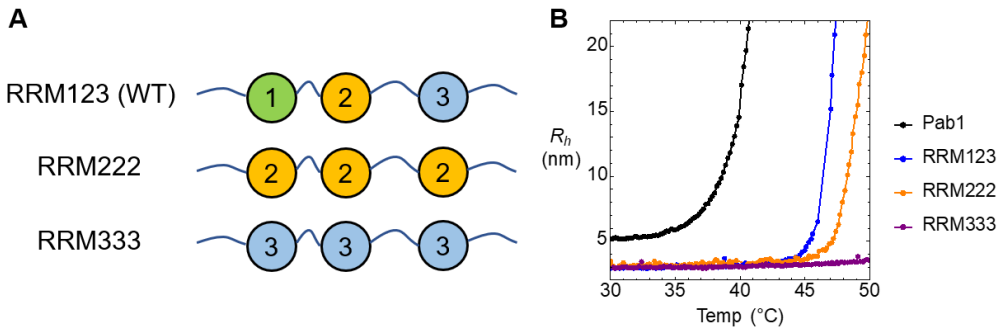


Figure 3.1: Triplet-RRM constructs. (A) Schematic of Triplet-RRM constructs. (B) DLS of RRM123 and triple-RRM constructs.

3.1.2 Unfolding of RRM1's helices help trigger condensation

We next asked whether the unfolding of the helices in RRM1 is a critical early step in the condensation process, as suggested by the HDX and DLS data. We made a 'bi-Cys' variant of RRM123 with a stabilizing disulfide bond inserted between the two helices on RRM1 (I61C, K94C) and lacking the two endogenous cysteines (Figure 3.2). Under reducing conditions where the disulfide bond was broken (1 mM dithiothreitol, DTT), the T_{demix} was considerably lower than that under non-reducing condition, 39 versus 48 °C. As a control, T_{demix} of WT RRM123 was found to have no response to DTT. The WT RRM123's T_{demix} is noticeably higher (46 °C) than the reduced bi-Cys version presumably because the cysteines are imperfect substitutions. Regardless, the large increase in T_{demix} upon formation of the disulfide bond indicates unfolding of one or both of RRM1's helices is necessary for triggering

condensation in RRM123 at T < 48 °C.

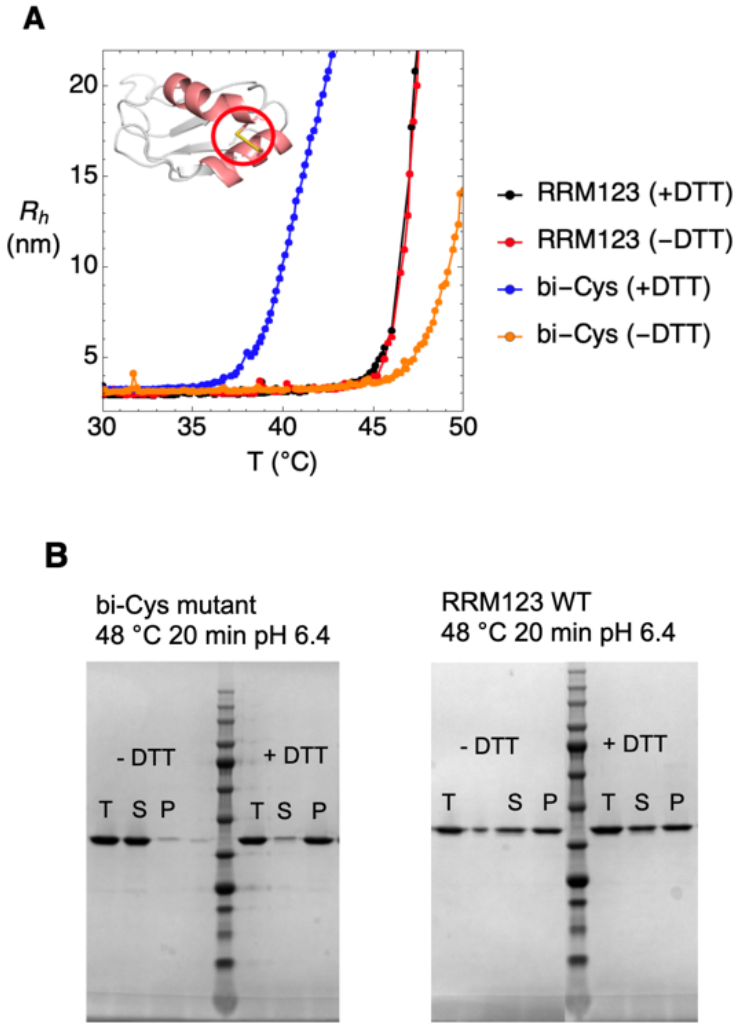


Figure 3.2: The condensation of bi-Cys mutant has big shift upon reduction by DTT whereas WT RRM123 has no response to DTT. (A) T_{demix} of bi-Cys is 10 °C earlier when 1 mM DTT is present. (B) TSP experiment yields the same result.

In addition, the bi-Cys mutant under oxidizing condition has nearly the same T_{demix} as RRM222 (Figure 3.3). This suggests that once RRM1's activation is inhibited, the onset of condensation is determined by RRM2, the next most activatable domain. This measurement illustrates how Pab1's RRMs can have their own activation temperatures and differentially influence condensation. The slope of the curve of bi-Cys was less than that of RRM222, presumably due to the difference of valency which affects the growth rate.

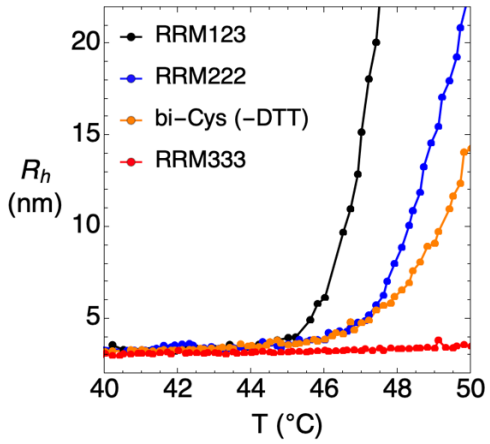


Figure 3.3: Bi-Cys mutant under oxidizing condition has nearly the same T_{demix} as RRM222.

3.1.3 Co-demixing experiments to test for stringency of activation and specificity

To evaluate whether each molecule must contain at least one activated RRM to participate in the condensation process, we examined whether triplet RRM constructs would co-demix with Pab1 at a temperature above Pab1's T_{demix} but below the triplet's T_{demix} (Figure 3.4). Specifically, we measured the extent of co-demixing of full-length Pab1 (T_{demix} = 39 °C) with RRM123, RRM222 and RRM333 (T_{demix} = 46, 48 and >50 °C, respectively) after incubation at 42 °C for 20 min or 46 °C for 10 min. The amount of material that co-demixed was identified using SDS-PAGE from amount in the pellet relative to the supernatant after centrifugation at 158,000 g for 10 min.

When the triplets were mixed with Pab1 at 42 °C, there was little (RRM123, 222) to no measurable (RRM333) triplet in the condensate. At 46 °C (at or below T_{demix} of the triplets), the majority of (>50%) RRM123 and almost all (>90%) RRM333 molecules remained in the supernatant. The majority of the RRM222 molecules, however, co-demixed, which we attribute to RRM222 having all three domains on the verge of activation, which

increases its binding valency and promotes condensation. Given the difference between the extent of co-demixing at 42 and 46 °C, and the lack of co-demixing for RRM333, we infer that domain activation is the major requirement for condensation. This inference implies that quinary interactions are insufficient to recruit unactivated (i.e., folded) RRM333 molecules into the condensate.

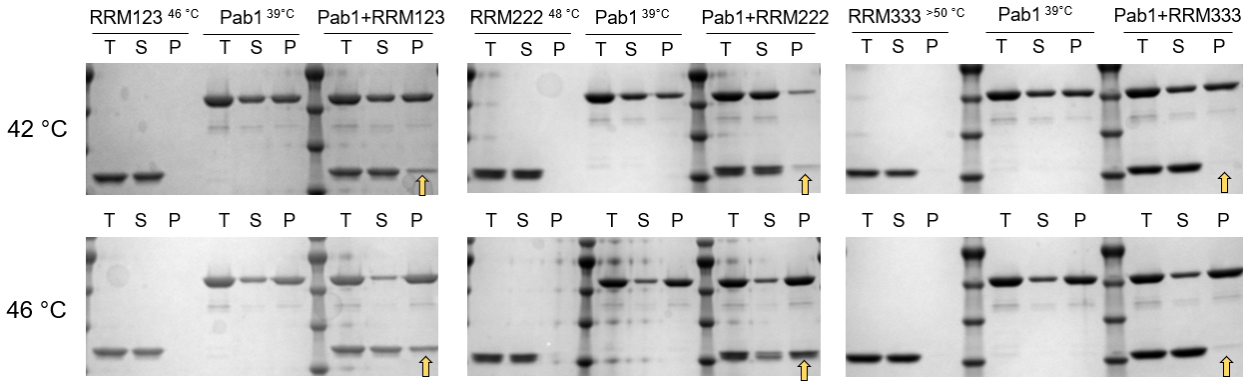


Figure 3.4: RRM123, RRM222 or RRM333 co-demixing with Pab1 at 42 °C (upper panel) or 46 °C (lower panel), both at pH 6.4. Yellow arrow indicates (expected) co-demixing band position. T_{demix} of constructs are provided (superscript).

To determine whether the RRM interactions are homotypic, we tested for co-demixing with a pair of constructs with orthogonal RRM composition: RRM113 ($T_{demix} < 39$ °C) and RRM222 ($T_{demix} = 48$ °C). When a mixture of RRM113 and RRM222 was heated at either 42 or 46 °C, the condensate was depleted of RRM113 but contained a small fraction of RRM222. (Figure 3.5). This co-demixing result indicates that heterotypic interactions can occur between RRM113 and RRM222, but in a manner that inhibits RRM113's condensation for this combination of triplets.

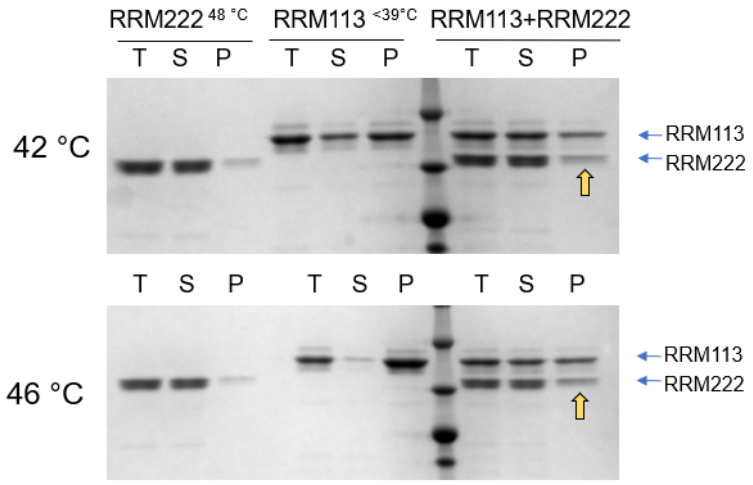


Figure 3.5: RRM222 co-demixing with RRM113 at 42 °C (upper panel) or 46 °C (lower panel). Co-demixing band position is indicated with yellow arrow.

Co-demixing experiments show that Pab1's condensation is distinguished from the IDR-dependent model of condensation that it is not merely controlled by quinary interactions, but necessitates the unfolding of RRM, which is determined by RRMs' intrinsic thermodynamic property. Furthermore, the co-demixing between RRM113 and RRM222 suggests that RRMs can form non-specific, heterotypic interactions, though there is an inhibitive effect that is not fully understood.

3.1.4 Sequential activation model with "thermodynamic specificity"

Based on the results discussed above including the RRM activation requirement, we propose a "sequential activation" model for Pab1's condensation (Figure 3.6). As temperature is increased, each RRM becomes sequentially activated reflecting the RRMs proclivity to partially unfold. NMR studies found that isolated RRM domains remained folded at temperatures above Pab1's T_{demix}. Therefore, activated RRMs need additional interactions with other regions to remain partially unfolded. The extensive heterogeneity in the HDX data, where populations with distinct conformations were observed, suggests a high degree of structural diversity with a heterogeneous set of contacts in the condensates. In this het-

erogenous milieu, intermolecular RRM interactions are facilitated by an increase in local concentration after the initial partnering of activated RRMs. This increase promotes the formation of other contacts and a highly interconnected network. Hence, a second stage of maturation may exist wherein the chains optimize their interactions and other weak interactions further stabilize the network [51, 43].

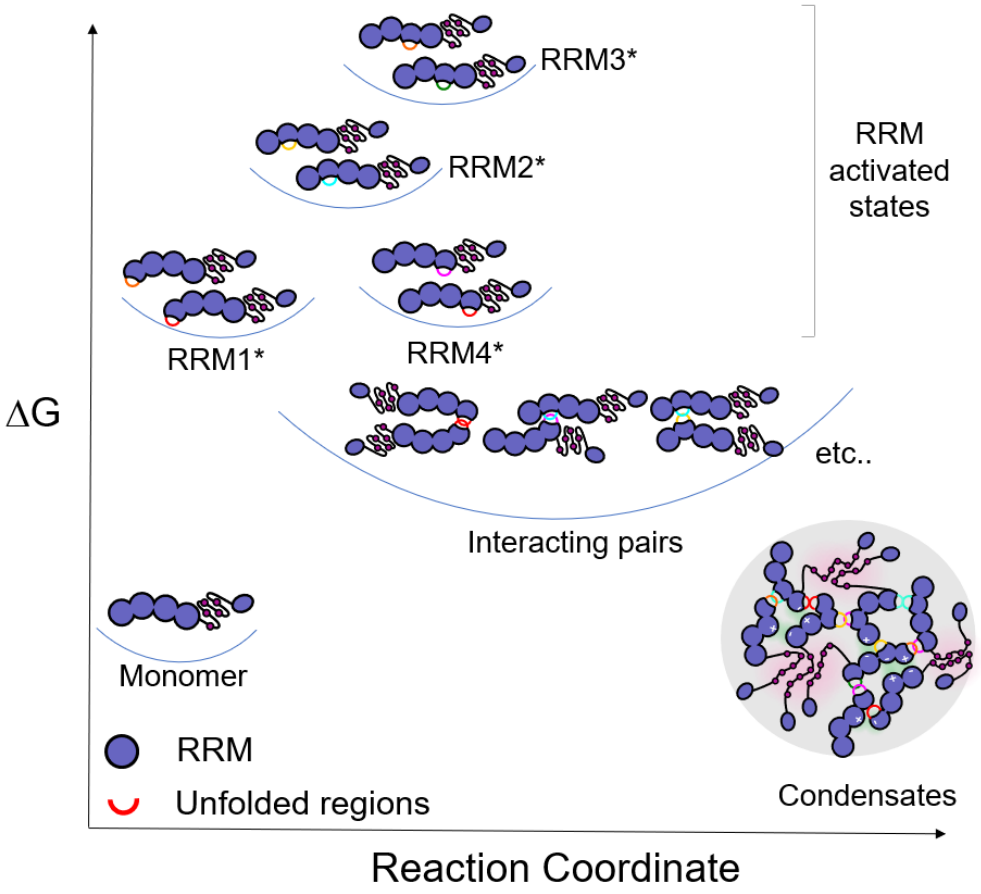


Figure 3.6: Sequential activation model for Pab1's condensation.

Because the RRMs have different activation temperatures and the ability to form stable contacts requires domain activation, the RRMs will tend to form contacts primarily with other RRMs having similar activation temperatures. We term this partner selection *thermodynamic specificity*. That is, condensation does not depend upon the presence of the highly complementary interface that one expects for biomolecular binding. Rather, quinary

interactions between activated RRM domains define the contact network. On a chain with multiple classes of RRM, the activation of each domain increases the valency and hence, the level of crosslinking should increase with temperature and lead to an increase in condensate stability and morphology, in principle, successively changing from dimers to chains and then to 3D networks.

3.1.5 The effect of urea on Pab1's condensation

Pab1's condensation is promoted by low concentrations of denaturant but blocked at higher concentrations, where 0.1 M urea reduced T_{demix} by 2 °C and 1 M urea abolished condensation (Figure 3.7). The contrasting effects of denaturant support our activation model where Pab1's condensation involves an initial partial unfolding step followed by the formation of a network of interchain interactions. The unfolding step is promoted by denaturant but the subsequent, condensate supporting interactions are hindered, with the net effect favoring condensation only at low urea concentrations.

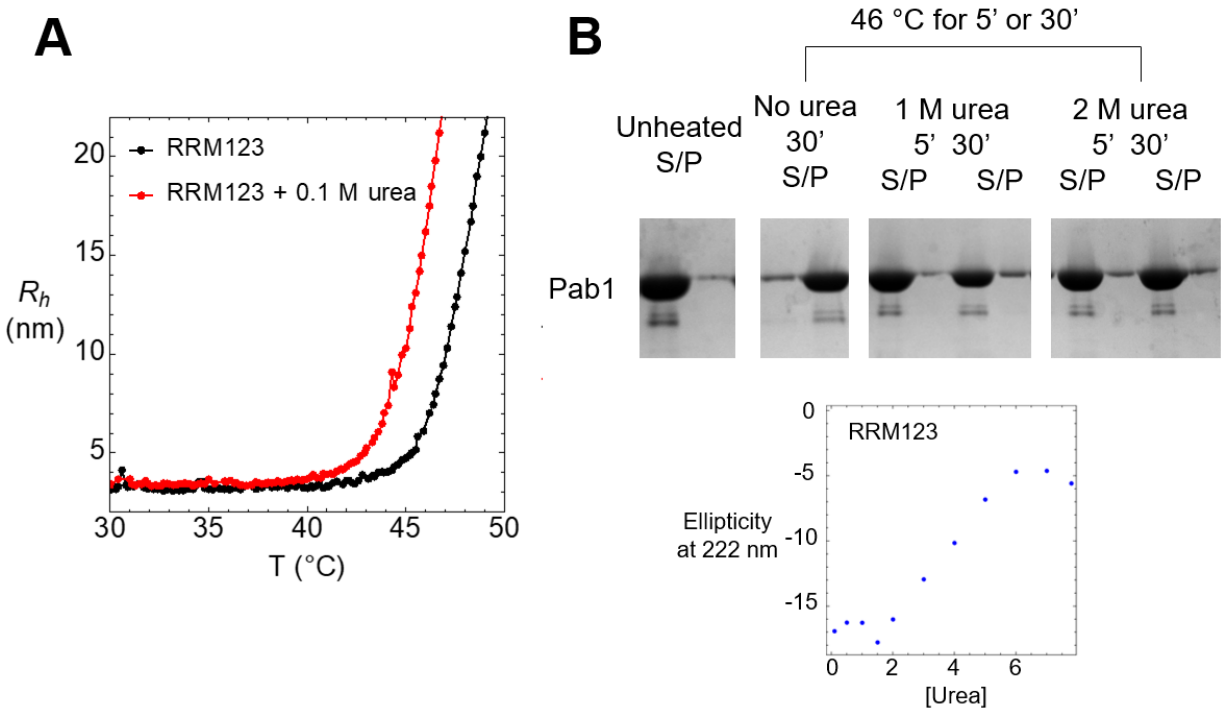


Figure 3.7: Urea's effect on Pab1's condensation. (A) Pab1's condensation with 0.1 M urea, measured by DLS. (B) Pab1's condensation is inhibited by 1 M urea, analyzed by TSP. Circular dichroism of RRM123 indicates no loss of secondary structure with 1 M urea.

3.2 Discussion

By triplet-RRM constructs we decipher the different participations in condensation of each RRM. Specially, RRM3 is the least involved RRM, presumably due to its high stability. RRMs have different activation temperatures, and co-demixing experiments show that the activation of RRM is required for its participation in condensation. In other words, Pab1's condensation is controlled by the partial unfolding of RRM instead of being completely dependent on quinary interactions. With the partial unfolding and different activation temperatures of the RRMs, the Pab1 condensation process is richer than classic condensation.

We propose a "sequential activation" model for Pab1's condensation. While the interactions between activated RRMs are likely non-specific, the different activation temperatures of RRMs unfolding imply a thermodynamic specificity on top of the quinary features of

biomolecular condensates, which is that an activated RRM is more likely to interact with other activated RRMs rather than inactivated ones. Whereas RRMs are similar enough to condense together, their thermodynamic difference leads to different thermal sensitivities and participations in condensation, which all together result in the unique temperature response of Pab1. The thermodynamic specificity is a centrist between stringent chemical specificity determined by molecular complementarity and non-specific quinary features, which endows Pab1 with unique sensitivity to transduce environmental stress signal and produce condensation.

3.3 Methods

DLS

DLS measurements were done as described in [57]. Measurements were performed in DynaPro NanoStar (Wyatt Technology). Each time point was averaged from five 6 s acquisitions filtering out samples with a baseline higher than 1.003 and analyzed in the DYNAMICS software with a cumulant fit to the autocorrelation function. Measurements were performed as a slow temperature ramp at 0.25 °C min_{-1} continuously from 25 °C or 30 °C. All experiments, unless noted, were performed at 15 μM protein in 20 mM HEPES, pH 6.4 with 150 mM KCl, 2.5 mM $MgCl_2$, and either in the absence or presence of 1 mM DTT. Samples were centrifuged for 30 min at 15,800 g at 20 C before DLS experiments. For measurements below pH 6, concentrated protein stocks in HEPES at pH 6.4 were diluted into 50 mM sodium acetate, 150 mM KCl, buffered to the indicated pH and samples were spun at 4 °C, 15,800 g for 5 min before DLS experiments.

TSP and co-demixing experiment

TSP protocol was adapted from [57] with modification. Concentrated protein stock were diluted into buffer containing 20 mM HEPES, pH 6.4 with 150 mM KCl, 2.5 mM $MgCl_2$

to a final concentration of 10 μM. For co-demixing experiments, both proteins are at a concentration of 10 μM. Sample was incubated at 42 °C for 20 min or 46 °C for 10 min followed by a centrifugation at 15,800 g for 10 minutes at 10 °C. Supernatant was collected as the soluble (S) fraction sample. Buffer was added to the pellet and the sample was centrifuged again under the same condition. After removing the supernatant, the pellet was resuspended in buffer as the pellet (P) fraction sample. Total, soluble, and pellet fractions are analyzed by SDS-PAGE.

CHAPTER 4

OTHER MOLECULAR FACTORS IDENTIFIED THROUGH SAXS, FRET, SIMULATION, AND MUTATIONAL STUDIES

We employed different biophysical approaches in combination with mutational studies to study the various molecular factors that promote Pab1's condensation. We found that histidines confer part of Pab1's sensitivity to pH as a histidine knockout mutant showed reduced pH sensitivity. We also investigated the role of Pab1's charged linkers by small-angle X-ray scattering (SAXS) and FRET in combination with mutational studies and found that Pab1's linkers are flexible and unstructured. Furthermore, SAXS results imply that Pab1's condensation is not due to auto-inhibition. Finally, to gain an insight into the interaction between RRMs, we employed simulations which suggest that non-specific heterotypic RRM interactions underlie condensation.

The study of histidines' role was designed and proposed by Dr. Joshua A. Riback. Darren Kahan performed the experiments and analyzed the data. The simulational study was done with the help of Dr. Xiangda Peng.

4.1 Results

4.1.1 Histidines partly confer Pab1's sensitivity to pH

Because histidines pKa of 6-7 is close to pH_{demix}, we examined the effect of Pab1's highly conserved RRM histidines on condensation [57]. We measured T_{demix} across different pHs in a variety of histidine deletion variants, including a ΔH123 construct in which all RRM histidines are substituted with the next most common amino acid in the Pab1 sequence alignment (Figure 4.2B). Histidine deletion likely lowers Pab1's pI as no histidines will be titrated when pH is lowered (Figure 4.1).

Condensation of the ΔH123 construct had a reduced pH sensitivity (Figure 4.2A), with

T_{demix} decreasing by 12 °C when pH was reduced from 6.4 to 5.0, whereas WT Pab1 experienced a 20 °C decrease. This decrease in sensitivity implies that histidines are partly responsible for Pab1's response to pH drop. A few histidines on RRM are adjacent (< 8 Å) to arginine or lysine residues. Specially, the positive charges on H53 and K105 are estimated to be 3.9 Å apart, providing a notable Coulombic repulsion when the histidine is protonated. Upon acidification, those histidines become protonated and the destabilizing Coulombic repulsion may cause unfolding of the region and subsequent condensation. Nevertheless, the titration of histidine does not fully explain Pab1's pH response. RRM4 has no histidines while pH-induced condensates still show a similar HDX pattern and comparable level of unfolding to Temperature-induced condensates (Figure 2.12).

The predicted net charges of all the histidine mutants are highly correlated with their T_{demix} at different pHs (Figure 4.2B). It implies that for these histidine mutants, being more positively charged favors condensation in a manner independent of the charge distribution on the structure. Some other mutants (linker charge mutant K123M and K306M, RRM3 charge mutant K268N and K241Q, and non-charge mutant A240I) deviate below the line, suggesting that T_{demix} has more sensitivity to those mutations.

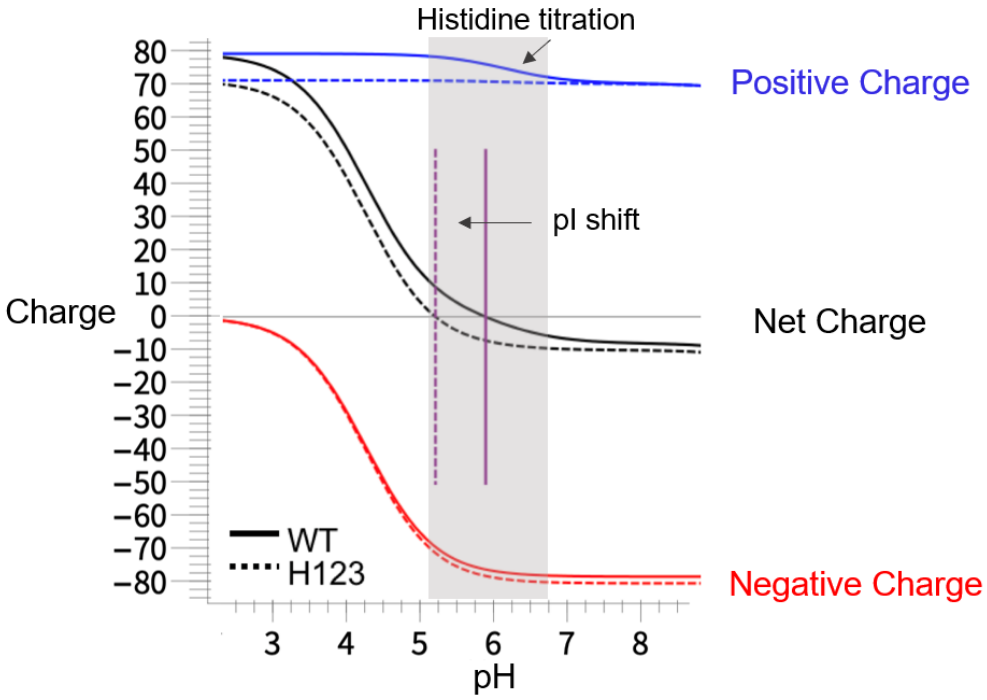


Figure 4.1: Titration curve of WT Pab1 and ΔH123 mutant. Histidine titration pH overlaps with Pab1's condensation pH range.

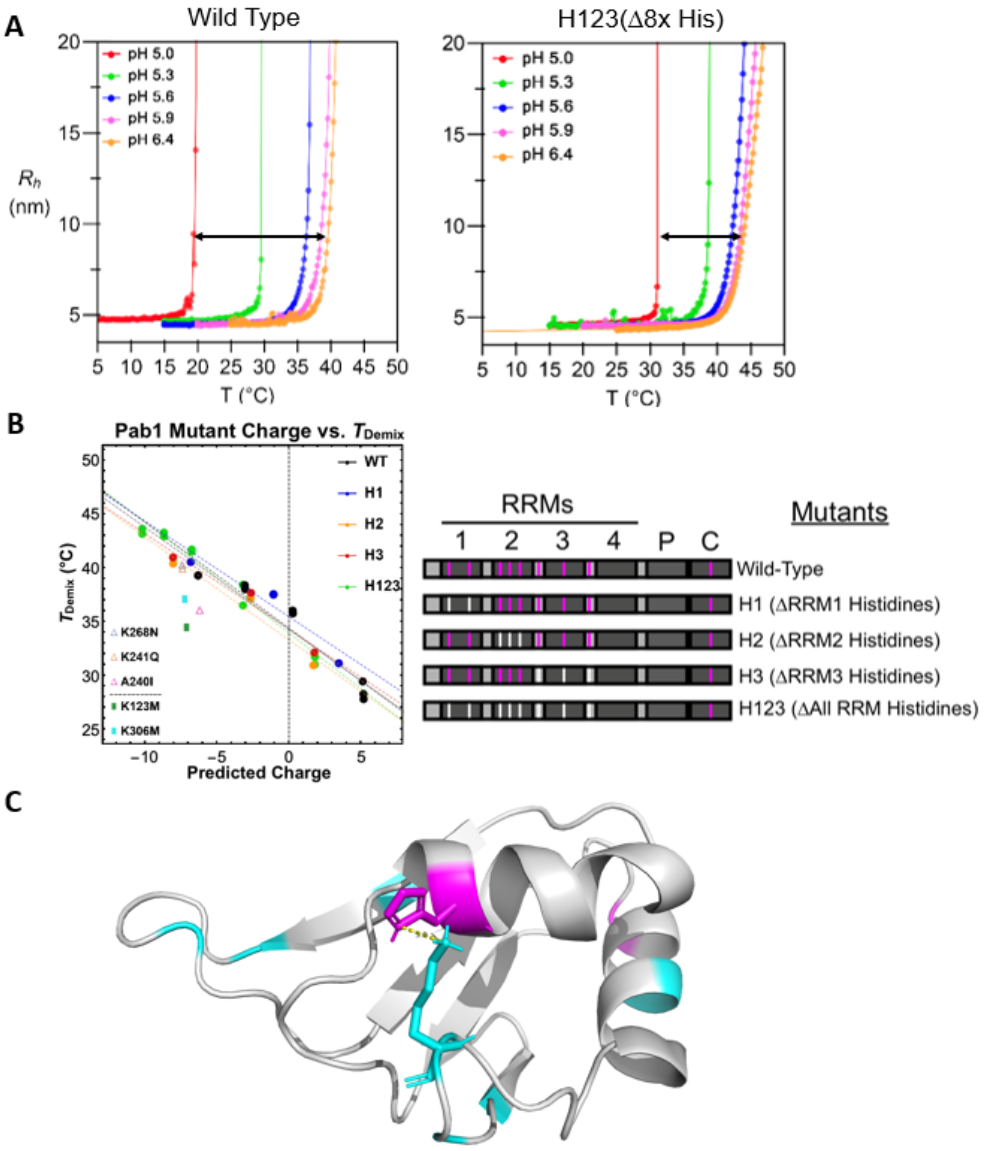


Figure 4.2: pH sensitivity of Pab1's condensation and Pab1's histidines. (A) DLS of H123 mutant and WT Pab1 at different pHs. (B) Predicted charges of Pab1's mutants correlate with their T_{demix} at different pHs. (C) Histidines (magenta) and other positive charge residues (Lysine and Arginine, Cyan) on RRM1. The positive charges of H53 and K105 (shown as sticks) are estimated to be 3.9 Å apart (yellow dotted line).

4.1.2 Pab1's condensation is irreversible and cannot be seeded

To investigate the role of nucleation, we performed a double temperature ramp experiment on Pab1. We heated the sample to the onset of condensation (at 40 °C, 1 °C above T_{demix})

and then cooled the sample to 25 °C for 20 min. The R_h level, 11.5 nm, did not change during cooling. We then reheated the sample and observed the same temperature-dependent profile in R_h as observed in the initial heat ramp (Figure 4.3). This finding indicates that the condensation process is not reversible by cooling and that nascent condensates do not nucleate condensation at temperatures below T_{demix}.

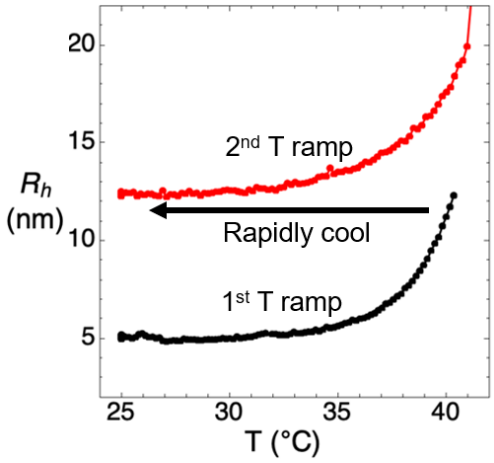


Figure 4.3: DLS experiment to test for nucleation with a 2nd temperature ramp after cooling.

Nevertheless, FRET studies with a 1:1 mixture of donor/acceptor single-labeled RRM123 (Alexa488/594, labeled on the same surface site of the ventral side of one RRM, Cys70 on RRM1, Cys158 on RRM2 or Cys251 on RRM3) yield kinetics reminiscent of a nucleation process (Figure 4.4D). After a ramp to and pre-equilibration right below T_{demix}, we started incubation slightly above T_{demix} and recorded FRET signal. We observed a brief lag phase suggesting that condensation is a nucleation process but only at temperatures above T_{demix}.

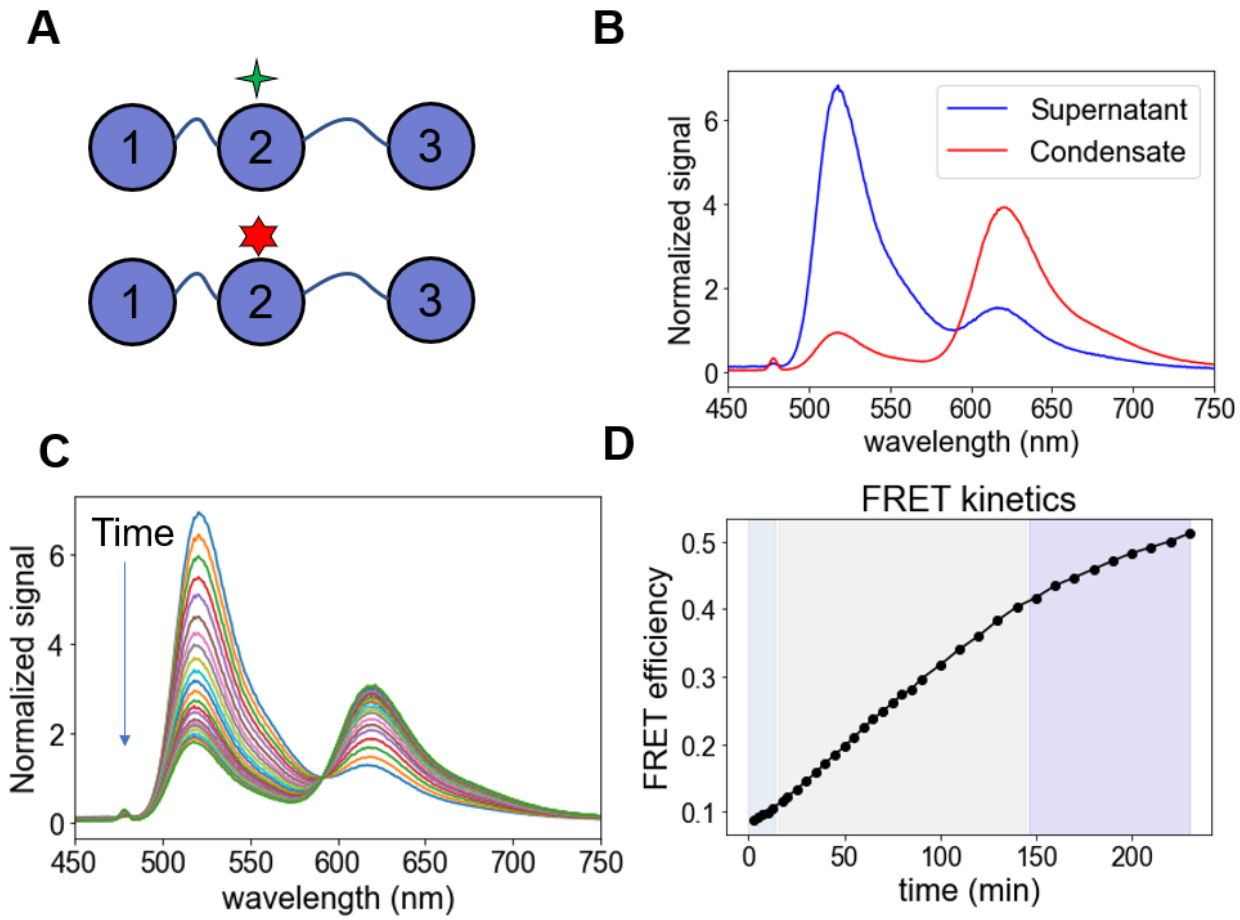


Figure 4.4: RRM123 condensation kinetics measured by FRET. (A) FRET labeling scheme. Either Alexa488 or Alexa594 is attached to a knock-in cysteine A158C on the ventral side of RRM2 through maleimide chemistry. (B) Higher FRET indicates associations between RRMs upon condensation. (C) FRET signal increases with incubation time. (D) FRET efficiency plotted against time where a lag phase is observed (light blue region), followed by a growth phase (light green), and a potential 3rd phase (purple).

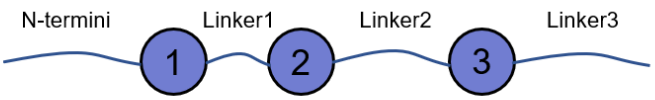
4.1.3 Test linkers' role and potential intramolecular inhibition

While most reported condensing protein systems are UCST [51], Pab1 exhibits an irreversible LCST condensing behavior [57]. Besides the unfolding of RRMs, we speculated that the thermal-responsive condensation may come from intramolecular inhibition that an initial event could be the unfolding of potential linker structure or the melting of any intramolecular binding interactions.

In the stickers-and-spacers model, spacers are not just connecting modules between stickers but can affect the phase behavior as they also mediate interactions augmenting the interaction network, though weaker than stickers' interactions by definition [11]. Another factor that one needs to consider is the excluded volume. A negative excluded volume corresponds to collapsed linkers while a positive excluded volume implies expanded linkers. The excluded volume of linker determines whether the protein is likely to form a network of droplet or gel [27], as it tunes the interplay between phase separation and percolation.

Pab1 has some linker regions connecting folded domains, including N- and C-termini, and interconnecting linkers between RRMs. In Pab1 structures from PDB structure database (PDB 6R5K, cryo-EM structure RRM1-4 in complex with polyadenylate RNA and Pan2-Pan3 deadenylase [59]), interconnecting linkers have helical regions rather than being completely unstructured. It leads to a potential model that the helical regions on linkers are unfolded upon temperature increase or pH drop, which provides flexibility for stacking of RRMs and triggers condensation. Note that the cryo-EM structure is Pab1 in complex with RNA and other proteins. Therefore, the helix does not necessarily exist in the solution structure of Pab1 alone.

The linkers between 4 RRMs are termed Linker1, Linker2, and Linker3, respectively. They are of different lengths: Linker1 is 9-residue, Linker2 is 15-residue, and Linker3 is 25-residue. Additionally, those linkers all contain a high percentage of charged residues, both positive and negative, depending on the linker (Figure 4.5). Previous study has elucidated an electrostatic basis of Pab1's association [57]. Therefore, we investigate both linkers' charge and structure by introducing perturbative mutations combined with other measurements including SAXS and FRET.



	N-termini 37 aa	Linker1 9 aa	Linker2 15 aa	Linker3 25 aa
Positive (R+K)	+2	+3	+4 (and 1 His)	+7 (and 1 His)
Negative (D+E)	-8	-1	-4	-3
Charged residues	10/37	4/9	8/15	10/25
Net Charge	-6	+2	0/+1	+4/+5

Figure 4.5: Charged residues on linkers.

Linker substitution

Two RRM123 mutants ('GSSG' and 'Scramble' mutants) were made where Linker1 and Linker2 were replaced with a Glycine-Serine linker or a scrambled-sequence respectively (Figure 4.6A). The scrambling is supposed to break any potential helical structure on the linker, making it an unstructured linker but maintaining the same amino residue composition. The two mutants both condense, suggesting that the two native linkers are not a strictly required component for the condensation. However, they both have T_{demix} about 3 degree earlier than WT RRM123, which implies an inhibitive effect of the native linkers (Figure 4.6B).

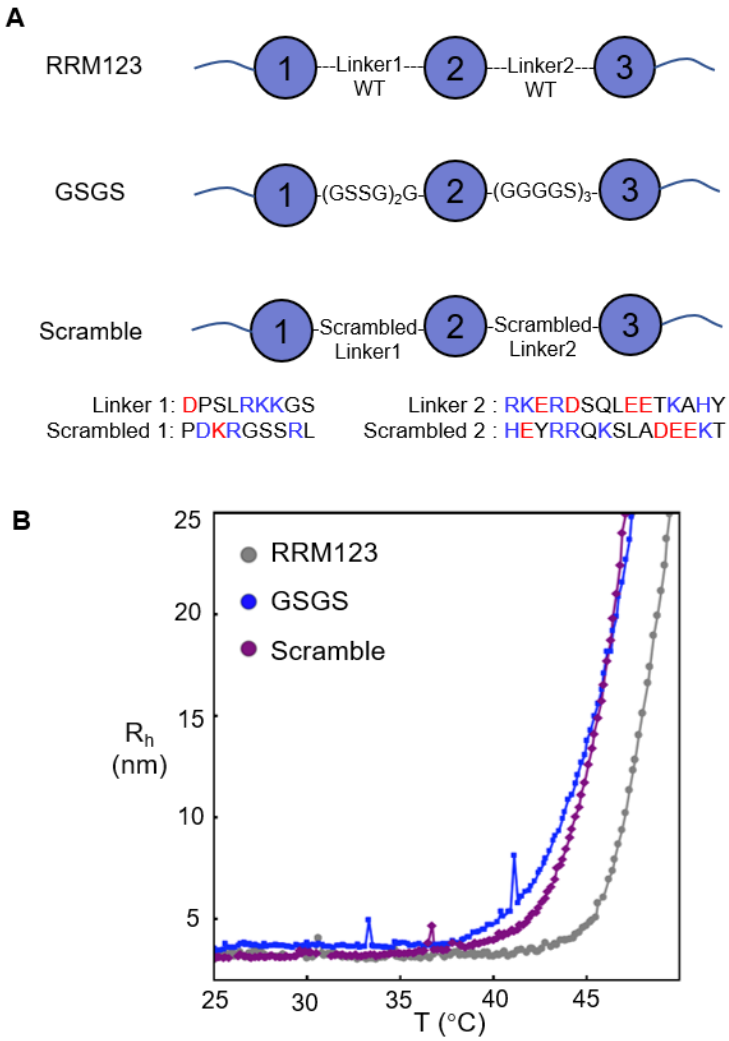


Figure 4.6: GSSG and Scramble mutants. (A) Linker1 and Linker2 are replaced with either GS linker or scrambled sequence linker with the same length while N-termini and Linker3 are kept the same as WT RRM123. Sequences of scrambled linkers are shown where charged residues are colored (blue for positive, red for negative). (B) DLS of WT RRM123, GSSG and Scramble mutants.

Urea dependence

The condensations of GSSG and Scramble mutants have minimal response to 0.1 M urea while WT RRM123's condensation is promoted by 0.1 M urea (T_{demix} decreased by 2 °C) (Figure 4.7).

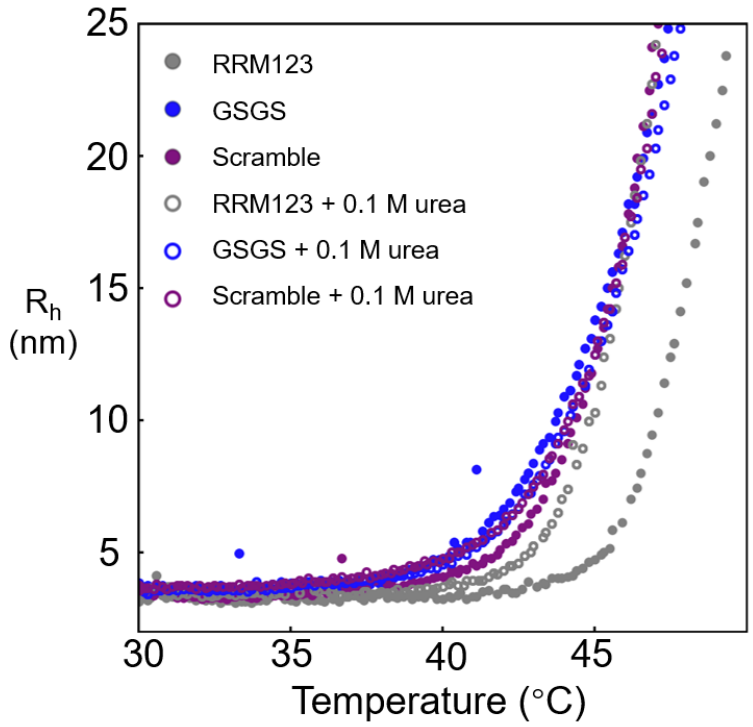


Figure 4.7: DLS of RRM123, GSGS and scramble mutant with and without 0.1 M urea. Data of measurement with 0.1 M urea are shown in open circles.

Glycine mutation (L120G, M301G)

Glycine was introduced into the helical region of either Linker1 or Linker3 which is supposed to be a helix-breaker and potentially facilitate condensation according to the linker-unfolding model. Instead, on DLS they show minimal T_{demix} difference from WT and even a slightly reduced rate in R_h growth (Figure 4.8A). The result is contrary to the linker-unfolding model.

Charge mutation (K123M, K306M)

It is known that Pab1's condensation is at least partially mediated by electrostatic interactions. The three linkers are highly charged (>= 40% charge residues, Figure 4.5). To investigate if linkers' charge plays a role in condensation, a charge-neutralizing lysine to

methionine mutation was introduced on either Linker1 or Linker3 (K123M and K306M mutant). Linker1 and 3 were chosen as they both have a positive net charge whereas Linker2's net charge is zero, so the K to M mutations reduce the net charge of Linker1 and 3. The two mutants both exhibit lowered T_{demix} corresponding to enhanced condensation, whereas K123M has a larger shift of T_{demix} (Figure 4.8B). This result confirms a participating role of linkers' charge in condensation. It is not fully understood that why the lysine is inhibitive on condensation, presumably resulting from the interactions the linkers form in the condensates.

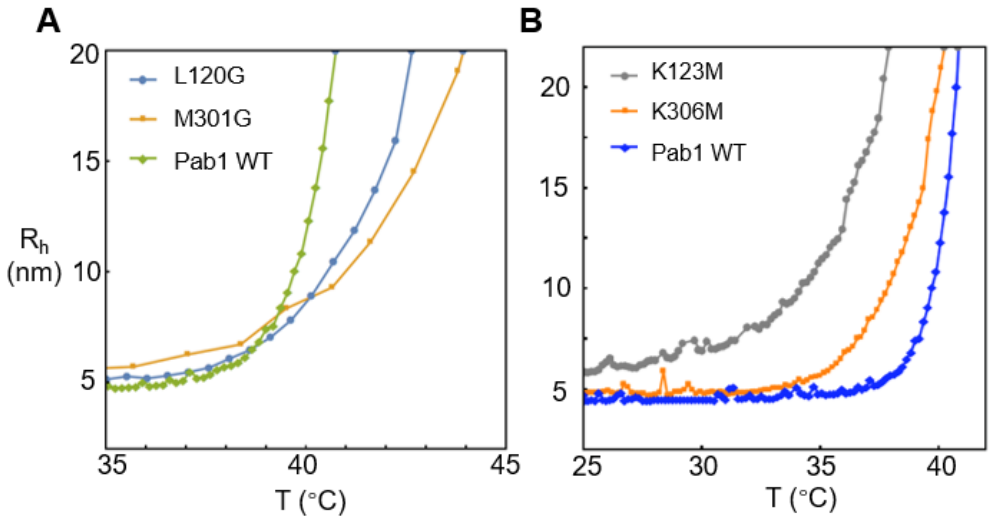


Figure 4.8: Linker glycine and charge mutants. (A) Glycine mutants L120G and M301G (B) Linker charge mutant K123M and K306M.

4.1.4 SAXS finds Pab1 to be extended and not auto-inhibited

To test if the GSSG and Scramble mutants' lower T_{demix} result from the unfolding of linkers, we conducted SAXS on the two mutants together with WT RRM123. Surprisingly, the three constructs have nearly identical pairwise distance distributions, P(r), and their R_g's are between 27.8 and 29.0 Å (Figure 4.9A). Based on the SAXS result, we simulated the conformation ensemble of RRM123 using our near-atomic level MD simulation algorithm Upside [32], where the structures of 3 RRMs are individually restrained, and temperature is

tuned to approximate the P(r) of experimental SAXS data. A simulated trajectory whose P(r) is close to the experimental P(r) shows that the conformation ensemble is expanded where linkers are flexible and RRMs are moving around with little restriction (Figure 4.9B).

We also made R12 (RRM1-RRM2) construct to test if the solution structure resembles the cryo-EM structure where RRM1 and RRM2 have a contacting interface and the Linker1 between 2 RRMs is collapsed (Figure 4.10A, bottom), which could be an intramolecular inhibition to the condensation. The SAXS data shows that R12 adopts a more expanded structure (R_g = 20.2 ± 0.6 Å, r_{max} = 75 Å) compared to Cryo-EM structure (R_g = 16.3 Å, r_{max} = 50 Å) (Figure 4.10A). The difference indicates that RRM1 and RRM2 are not as close in solution as in the Cryo-EM structure. We again utilized Upside to simulate a conformational ensemble with comparable distance distribution to experimental data. To match the experimental P(r) requires an expanded conformational ensemble. The residue-residue contact map from the trajectory shows that RRM1 and RRM2 do not form stable contacts as in the cryo-EM structure (Figure 4.10B).

We ran SAXS measurements on R12 and R123 (RRM1-RRM2-RRM3 without terminal linker on both ends) with or without 0.5 M urea to further confirm the promotive effect of 0.1 M urea is not from the unfolding of linkers or breakage of any potential interaction between RRM1 and RRM2 that restricts Pab1 from forming intermolecular interactions. R12 exhibited a completely identical P(r) distribution and R123 showed a minimal difference where R_g increased by 1 Å (Figure 4.11). Therefore, 0.1 M urea is unlikely to promote Pab1's condensation by unfolding those interconnecting linkers but by lowering the overall stability of RRMs. It is unclear why GSSG and Scramble have little response to 0.1 M urea while SAXS does not support the presence of any structure in the native linkers.

Pab1's condensation is affected by ionic strength (Figure 4.12B and [57]). To test if there is any auto-inhibition arising from intramolecular electrostatic interactions between RRMs and charged linkers (Figure 4.12A), we conducted SAXS on Pab1 and R123 with different salt concentrations. Both Pab1 and R123 have the same P(r) profiles under 50 mM, 150

mM, or 350 mM KCl (Figure 4.12C). That high salt does not expand Pab1 implies that there is no electrostatic auto-inhibition.

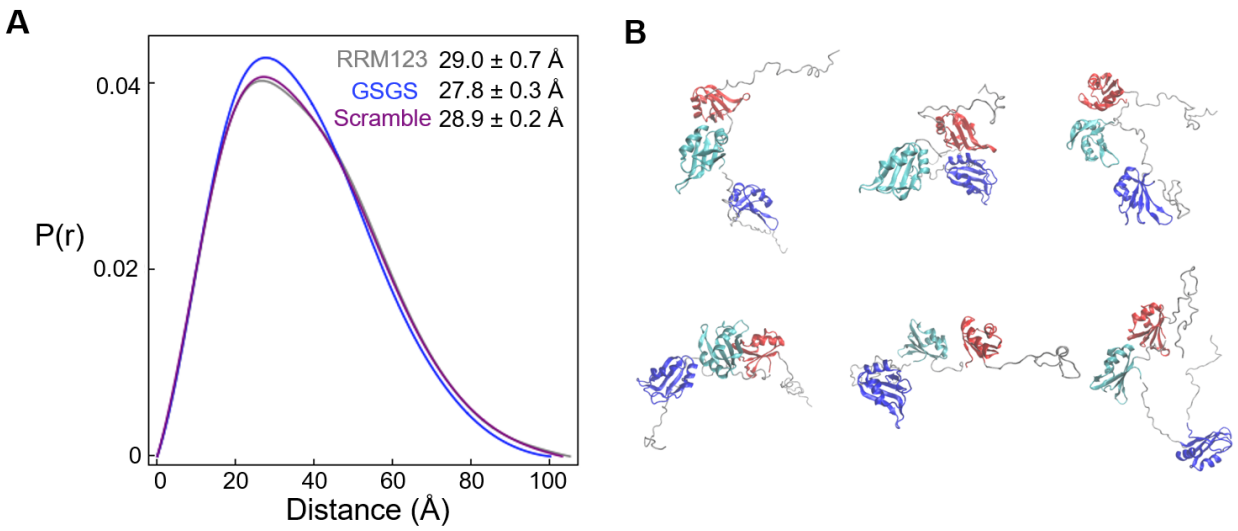


Figure 4.9: SAXS and simulations to study linkers' structure and Pab1's conformation. (A) Distance distribution for RRM123 WT, GSGS, and Scramble mutants (B) Simulated upside trajectory based on the experimental SAXS data shows that linkers are flexible and RRM123 is adopting a highly heterogeneous conformational ensemble.

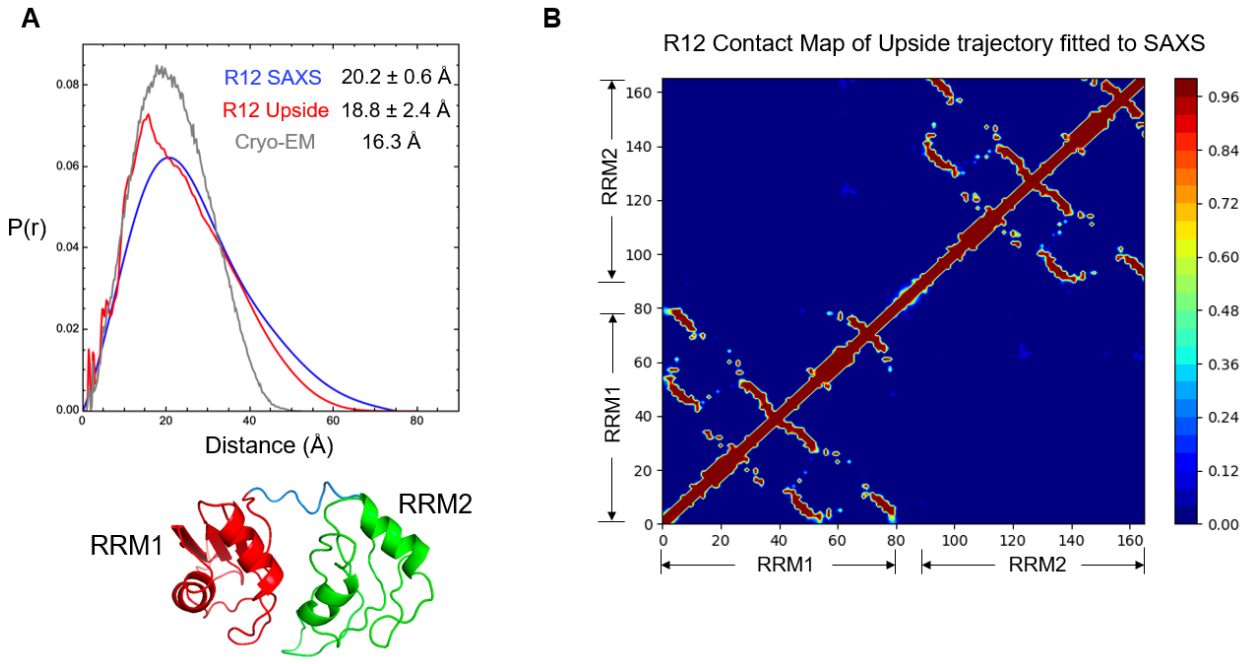


Figure 4.10: SAXS, Cryo-EM, and Upside simulation of R12 structure. (A) SAXS profile of R12 compared to Cryo-EM structure (6r5k.pdb). (B) Contact map between RRM1 and RRM2 obtained using Upside simulations at a Temperature chosen to match the experimental SAXS data.

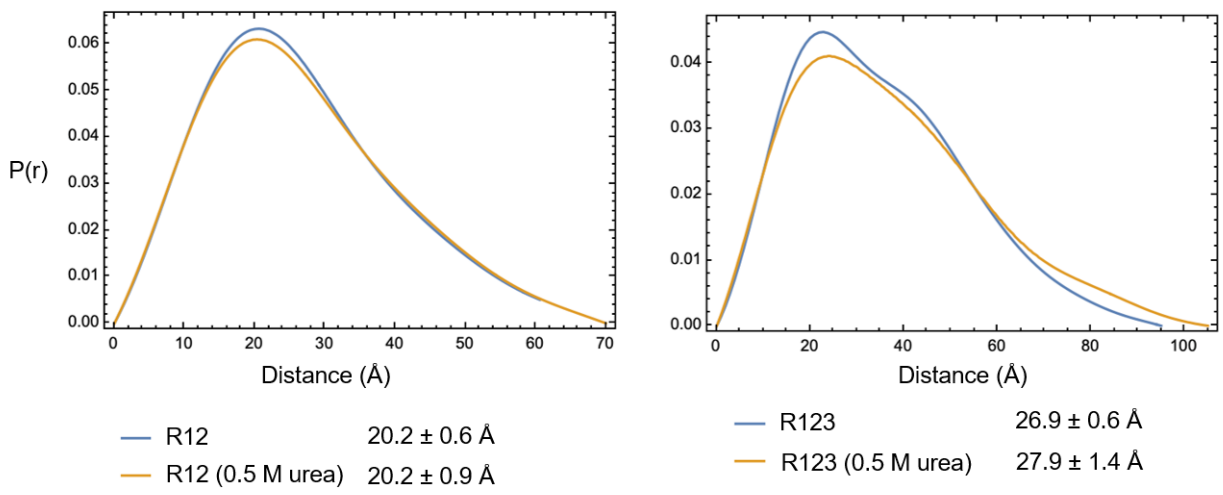


Figure 4.11: Distance distributions for R12 and R123 with or without 0.5 M urea.

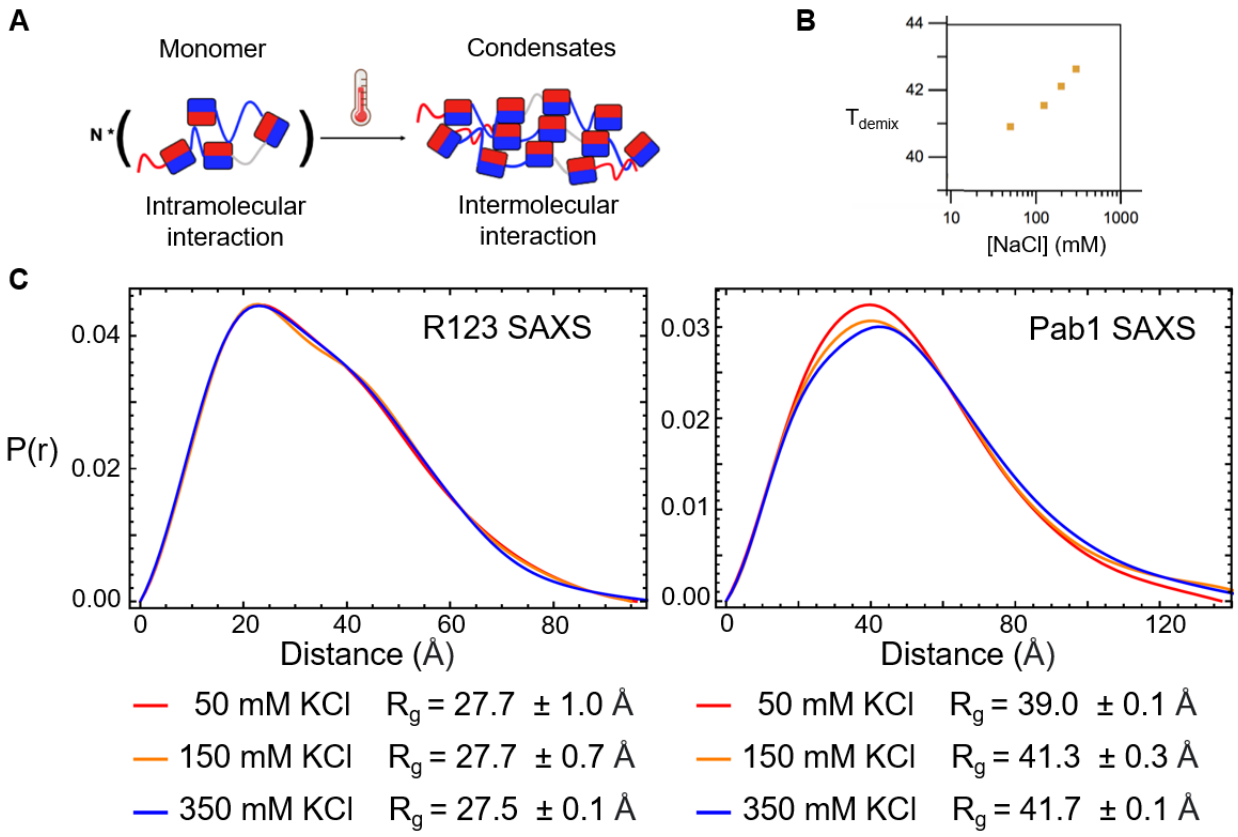


Figure 4.12: Test the auto-inhibition for Pab1's condensation. (A) Schematic of a auto-inhibited case where intramolecular interactions prevent Pab1 to form an intermolecular network through the stacking of charged-surfaces of RRMs. RRM's two differently charged surfaces and linker charges are indicated by color (red for negative charge, blue for positive charge). (B) Salt concentration influences T_{demix}. (C) SAXS profile of R123 and Pab1 at different salt concentrations.

4.1.5 FRET shows that there is no conformational expansion upon condensation

We took advantage of the benefit of FRET that it can probe within the condensates and provide proximity information to investigate if Pab1 has any significant conformational change upon condensation. We mixed Pab1 with Alexa 488 and Alexa 594 to label its two endogenous Cysteines (C70 on RRM1, C368 on RRM4. The third native cysteine on RRM1 C109 is substituted with an isoleucine.) (Figure 4.13). We did not use reverse phase chromatography to separate D-D, D-A and A-A labeled Pab1 as Pab1 might condense with acetonitrile and

it is unclear if Pab1 refolds to its native structure. Instead, we removed the unreacted free dyes and measured the heterogeneously labeled protein to track if there is any change in the average FRET level upon condensation. Upside simulations find that the distance between the two labeling sites fluctuates between 13 to 106 Å, which falls into the detectable range of the Alexa488-594 dye pair which has a Förster distance of 60 Å. Therefore, if there is an intramolecular inhibition, the closed to open conformational change should be detectable as a decrease in FRET signal.

We diluted the labeled Pab1 with unlabeled Pab1 to arrive at a final sample concentration of 10 μM where $\sim$10% molecules are labeled with dyes. The dilution with unlabeled protein reduces the probability that labeled molecules have other labeled neighbors so that the fluorescence primarily reflects intramolecular information. The sample was heated with a slow temperature ramp at 0.5 °C min^{-1} from 30 °C to 45 °C where most Pab1 molecules readily condense. The emission spectra exhibited minimal differences upon heating (Figure 4.13). This result implies that Pab1 has no significant large-scale conformational change upon condensation, which is consistent with the SAXS study that Pab1's condensation is not auto-inhibited.

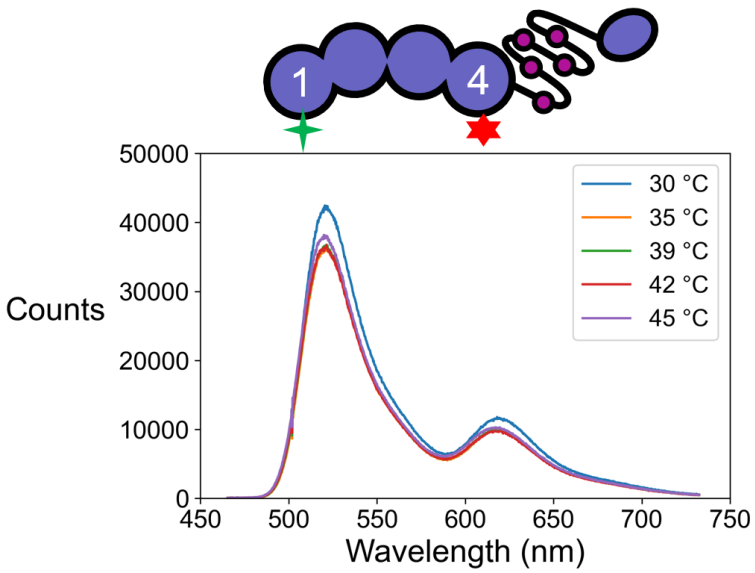


Figure 4.13: FRET on bi-labeled Pab1. Pab1 is randomly labeled with equal amounts of Alexa488 and Alexa594 dyes so that some Pab1 molecules have DD, AD and AA labels, as indicated with the green and red stars. A 10 μM Pab1 sample with 10% labeled are heated from 30 °C to 45 °C with spectra recorded.

4.1.6 Data of other Pab1 mutants

Truncated constructs: R123, R12, RRM123-

R12 and R123 constructs still have condensation-like behavior when heated at 46 °C that the protein samples turned turbid/cloudy and some fraction became pelletable. However, the DLS under the same temperature ramp does not produce the typical T-dependent curve where the R_h has a sharp increase at around T_{demix}. Rather, R12's R_h profile on DLS suggests a multimeric aggregated state at room temperature, which impedes the determination of T_{demix} by DLS. The R_h trace of R123 has two distinct phases: an initial growth starting at 36 °C with a slower profile compared to other Pab1 constructs, and at 47 °C the growth rate became faster (Figure 4.14B). The two temperatures are close to the activation temperature of RRM1 and RRM2 respectively. We hypothesize that the first phase corresponds to a condition where only RRM1 is activated and Pab1 can only form dimer or oligomerized chains but cannot grow into an extensive 3D multimer network due to the limitation of

valency (number of activated RRMs). Therefore, the first phase would yield a slow-growth profile. Consistently, previous DLS result [57] showed that the R_h of Pab1 increased slowly when incubated at 36 °C, whereas R_h remained unchanged below this temperature. At 47 °C, the activation of the second activatable RRM (i.e. RRM2), would explain the second phase of R123's R_h growth. The growth rate became much faster due to an increased valency of activated RRMs, which enabled formation of a 3D network with a larger R_h.

That R123 shows the two-phase growth behavior which is different from RRM123 suggests that the N-termini of RRM123 might be affecting RRM1's pairing with other molecules after activation. For example, when N-termini was truncated from RRM123 (RRM123- construct, RRM1-RRM2-RRM3-Linker2), similar to R123, its R_h went up at 35 °C, earlier than RRM123 and close to the expected activation temperature of RRM1, with a lifted baseline at ∼9 nm, suggesting a stable multimeric state in solution before condensation (Figure 4.14C). We speculate that the N-termini is inhibitive to RRM1's pairing with another RRM1 on other Pab1 molecules.

RRM213

RRM213 has a slightly lower T_{demix} than RRM123 (Figure 4.14B). While it can plausibly be argued that the interaction between RRM1 and RRM2, which is removed in RRM213, inhibits condensation, the SAXS data is inconsistent with this hypothesis. Presumably, the RRM213's enhancement arises from a directional effect related to the relative positions of 3 RRMs affecting their optimal arrangement in the condensates.

RNA-binding-defective Pab1-15

This mutant is from [12]. The four mutations (Y83V, F170V, F263V, F366V) target a conserved aromatic residue (Y/F) in the RNA-binding cleft and lead to an RNA-binding incompetent mutant. Pab1-15 still condenses and has an earlier T_{demix} (3 °C shift) compared to WT Pab1 (Figure 4.14C). This implies that Pab1's capability of condensation is

independent of its RNA-binding affinity. This DLS data is provided by Dr. Haneul Yoo.

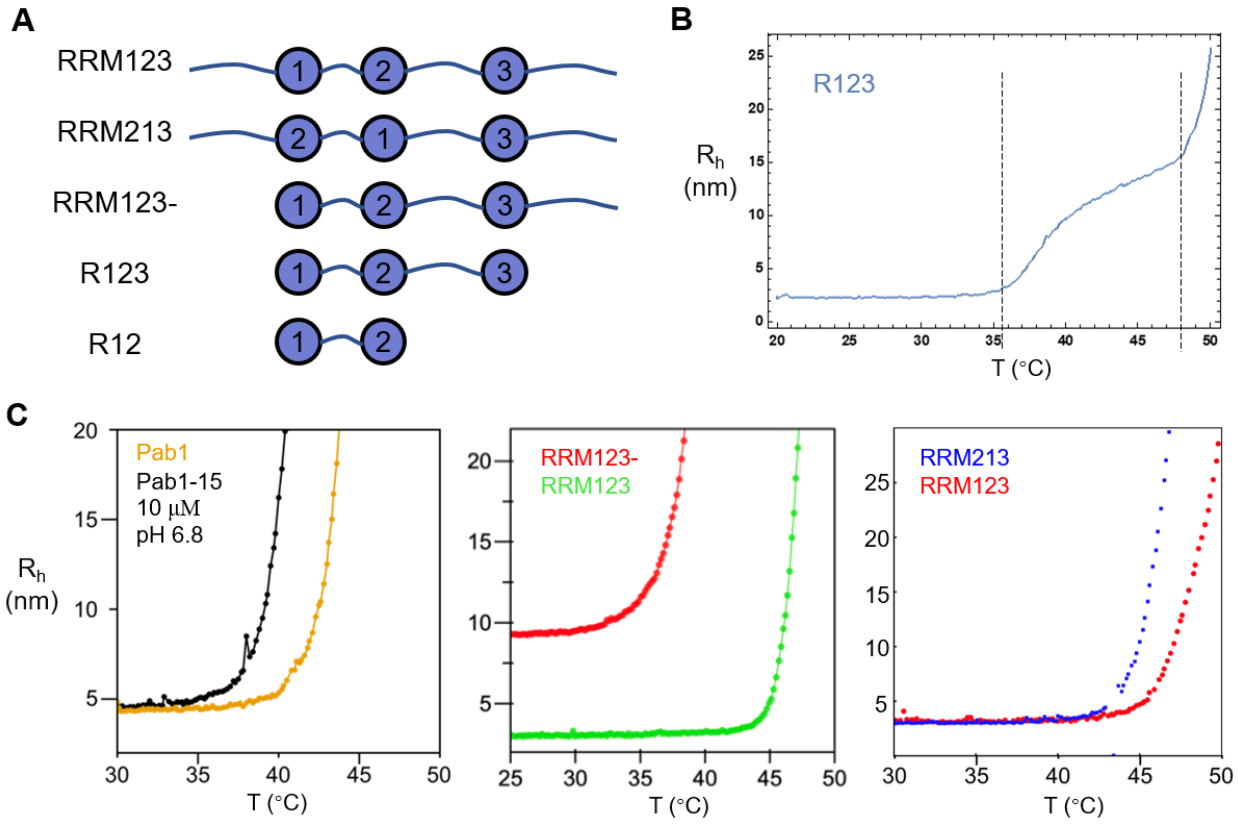


Figure 4.14: DLS of Pab1 mutants/constructs. (A) Schematic illustration of constructs' structure. (B) R123 has a different-shaped DLS curve with 2 distinct growth phases. (C) DLS of Pab1-15 (Figure courtesy of Dr. Haneul Yoo), RRM123- and RRM213 construct.

4.1.7 Simulation to study the interaction between RRMs

While co-demixing experiments suggest that RRM interactions are dependent on the activation step, and RRM113 and RRM222 can form heterotypic interactions, we do not have a clear picture of the interaction modes between the different RRMs. The inhibition of RRM222 on RRM113's condensation is also mysterious. In this section, we used a simple model where different interaction modes are tested for recapitulating the features of different constructs' T_{demix} and their co-demixing behavior.

Simulation set-up

We modeled RRM constructs as spheres connected by linkers. In our model, RRMs are core spheres with a 3 Å radius. A spring potential is added to penalize the overlap of spheres. Linkers are uniformly 4 Å long and for simplicity, and do not contribute to interactions (Figure 4.15).

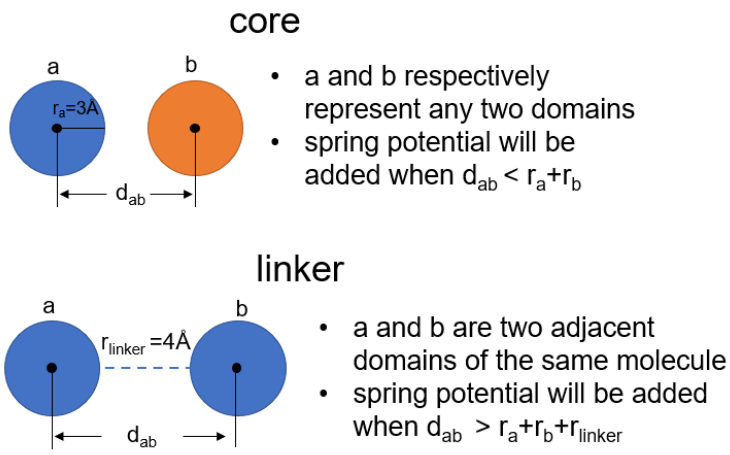


Figure 4.15: RRM constructs modeled as spheres connected by linkers.

Interaction between spheres is modeled as a sigmoid distance potential, and only inter-molecular interactions are included in the model (i.e. the spheres will not interact with another sphere on the same chain).

As we have found that RRM will be activated at different temperatures, in the model there are two types of interactions: activated and non-activated. Activated interaction has the same distance dependence as non-activated interaction, which is that the width of the sigmoid is the same; but the sigmoid of activated interaction has an increased depth corresponding to a higher strength interaction. The difference between active and non-activated, an extra energy from activation, is a temperature-dependent sigmoid to reflect the RRM activation at certain temperatures (Figure 4.16A). From the DLS data of triplet-RRM constructs, the activation temperatures of RRM1 < RRM2 < RRM3. Though we

do not know how RRM444 condenses, from HDX-MS where RRM4 has a similar or even higher level of unfolding compared to RRM1 and more unfolding than RRM2, a reasonable inference is that the activation temperature of RRM4 is lower than RRM2.

Note that even for non-activated RRMs, they will have the baseline-level non-activated interaction rather than being completely non-interacting. This is based on non-activated RRMs can still form weak quinary interactions such as electrostatics. After activation, the interaction becomes stronger as unfolded interactions start to participate. The values of the two types of interactions (1.8 k_BT for non-activated, 2.5 k_BT for activated) are not arbitrary but set in a way that non-activated interactions will not cause a collapse of the molecules whereas activated interactions are sufficient to.

An energy penalty was included when an RRM has more than one RRM from other molecules in close proximity (distance <= 10 Å) to ensure that an RRM will only interact with one another RRM. Simulations were run within a cavity of radius of 80 Å to simulate a constant concentration (Figure 4.16B).

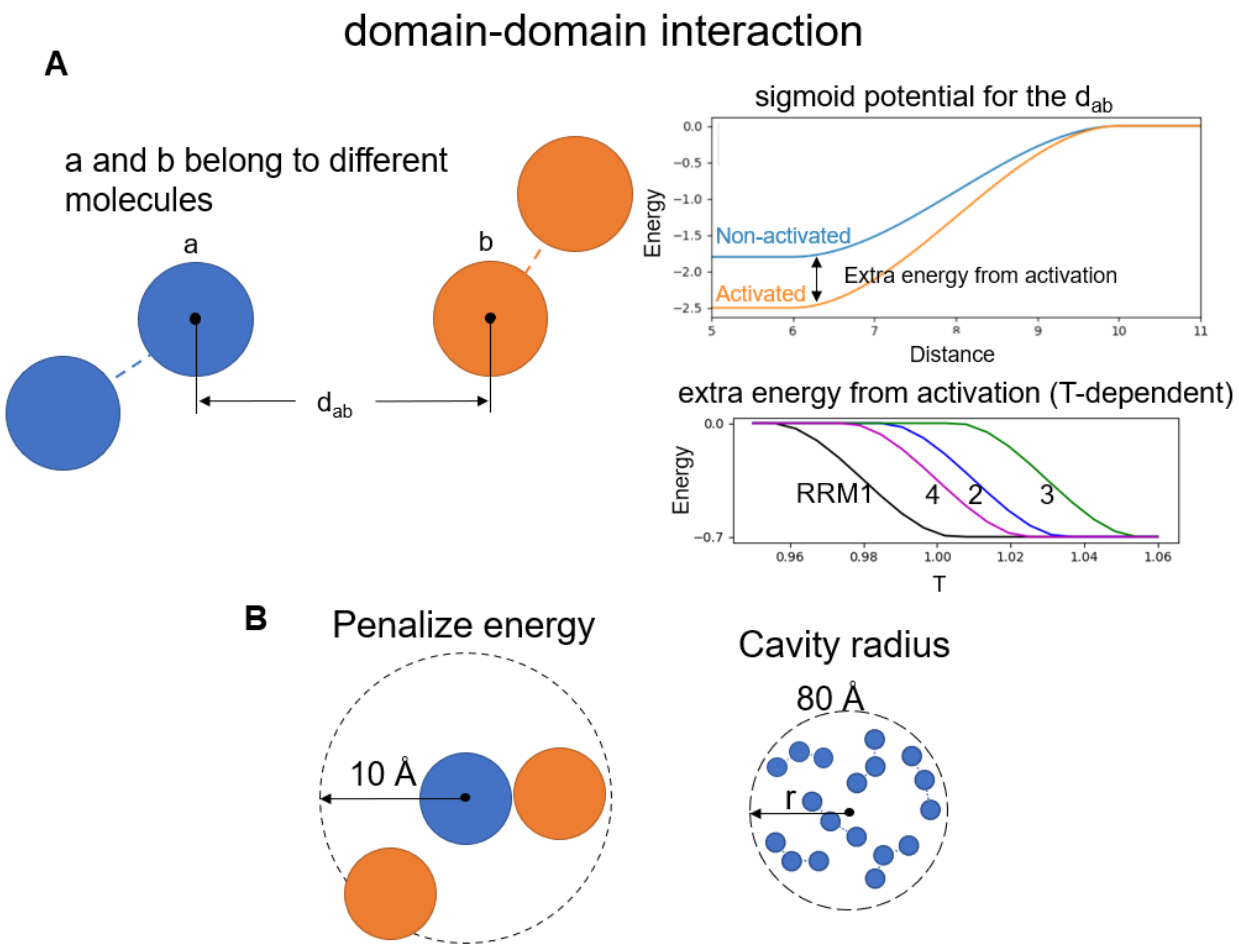


Figure 4.16: Modeling of domain-domain interactions. (A) domain-domain interactions are modeled as distance-dependent sigmoid, either non-activated or activated. The extra energy from activation is a temperature-dependent sigmoid where activation temperatures of RRM1 < RRM4 < RRM2 < RRM3. (B) energy penalty when more than one domain is in proximity.

Different RRM interaction modes

There are two questions regarding RRM interaction modes:

1. Are RRM interactions specific? Specificity refers to a case where RRMs form active homotypic interactions only: RRM1-1, 2-2, 3-3, 4-4. For non-specific scenario heterotypic interactions will be included: RRM1-2, 1-3, 2-3, etc.

2. Do both RRMs need to be activated or only one RRM does? e.g. At a temperature $T_1 < T < T_2$, (T_1 and T_2 are the activation temperatures of RRM1 and RRM2 respectively),

will RRM1*-2 (* represents an activated RRM) form strong interaction?

Based on those two questions, we test 3 different scenarios with decreasing stringency of interaction:

1. Only homotypic interaction will become activated (more stabilizing) above the activation temperature.

2. Heterotypic interactions when both RRMs are activated will become more stabilizing.

3. Heterotypic interactions when at least one RRM is activated will become more stabilizing.

Note that the only input for the model is that 4 RRMs have different activation temperatures $T_1 < T_4 < T_2 < T_3$. We are testing what RRM interaction mode recapitulates the experimental data including the T_{demix} of different RRM constructs (RRM111 < RRM123 < RRM222 < RRM333), and co-demixing features. A decrease of the R_g denotes a collapse of the system, which is analogous to condensation.

Simulation results compared to experimental data

Under scenario 1 where only homotypic interaction will become more stabilizing above the activation temperature ($T_1 = 0.98$, $T_2 = 1.01$, $T_3 = 1.03$), RRM111 and RRM222 collapse around the activation temperature of their composing RRM, consistent with the data that T_{demix} of RRM111 is less than that of RRM222. However, RRM123 does not readily collapse before RRM3's activation temperature, which is in contrast to the experimental result where T_{demix} of RRM123 is lower than RRM222 (Figure 4.17). This trend holds regardless of the strength of activated interaction.

The contradiction is that for RRM123 whose 3 RRMs are activated at different temperatures, in this case where only homotypic interaction is contributive, RRM123 is less favored to condense compared to RRM222 when RRM3 is not activated (at $T < T_3$), as the valency is less by 1, even though its RRM1 will be activated the earliest. But experimental data that RRM123 condenses earlier implies that having RRM1 is advantageous for condensation so

that RRM123 condenses before RRM2's activation temperature. Therefore, this scenario is not consistent with experiment.

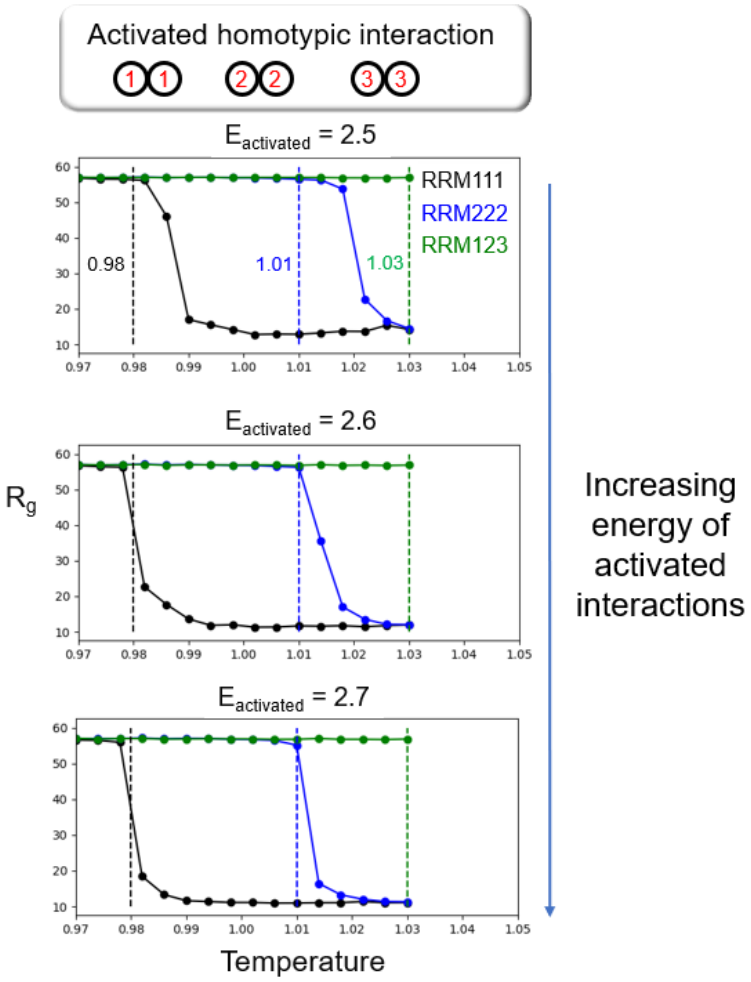


Figure 4.17: Scenario 1 with only activated homotypic interactions above the corresponding activation temperatures.

Under scenario 2 where heterotypic interactions will become more stabilizing when both RRMs are activated, the same contradiction with experimental data occurs that RRM123 is condensing later than RRM222 (Figure 4.18). When stronger heterotypic interactions require both RRMs to be activated (i.e., at the higher RRM activation temperature), the condensation is dominated by the RRM with the highest activation temperature. So RRM123 will condense less easily as it has an inactive RRM3, which is contrary to experimental result.

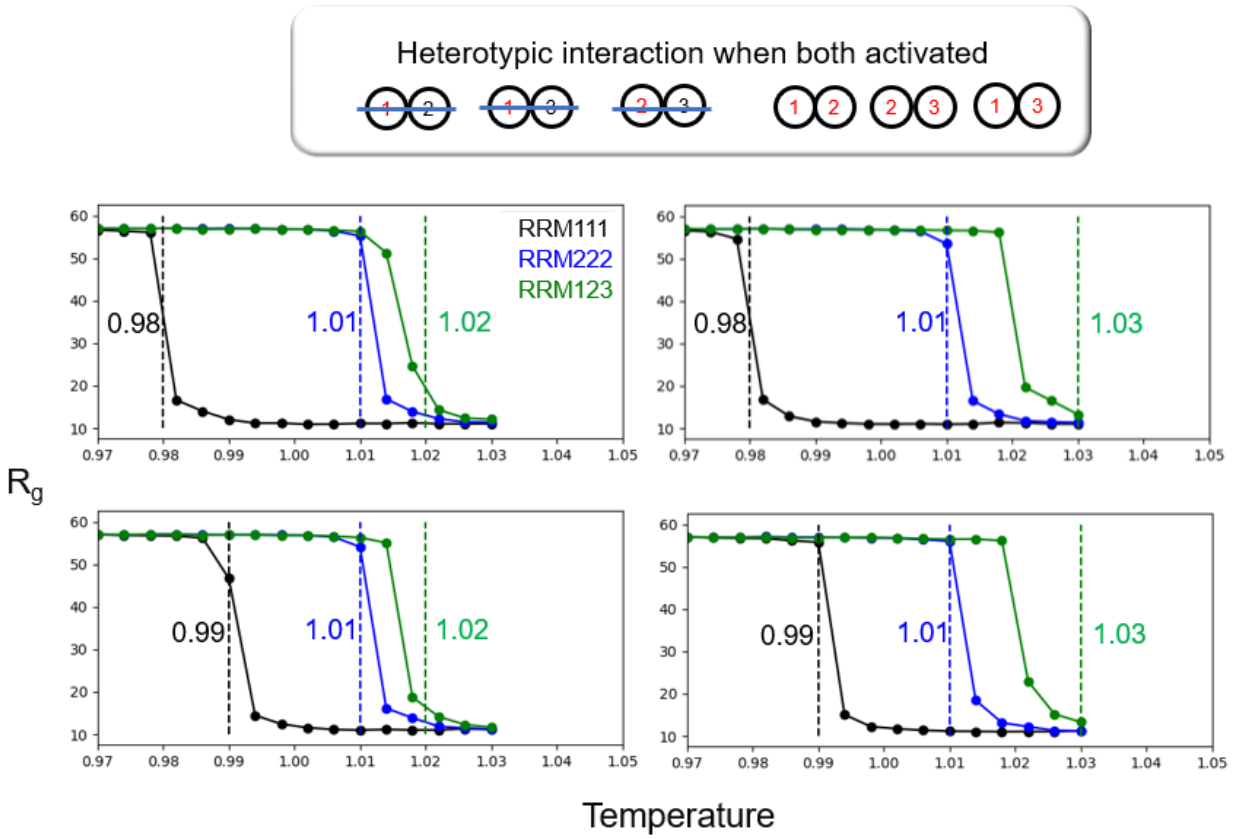


Figure 4.18: Scenario 2 with active heterotypic interactions when both RRMs are activated. Different combinations of T_1 and T_3 are tested while keeping $T_1 < T_2 < T_3$.

For scenario 3 where heterotypic interactions will become more stabilizing when at least one RRM is activated (i.e., above the lower RRM activation temperature), we set that when $T >= T_A + \frac{1}{4} * (T_B - T_A)$ for $T_A < T_B$, which is a quarter of the way between the two activation temperatures, the interaction between A and B goes to the activated energy level (Figure 4.19A). The simulation result successfully reproduced the order of T_{demix} that RRM111 < Pab1 < RRM123 < RRM222 < RRM333 (Pab1 is simulated as RRM1234) (Figure 4.19B). It is presumably because the stronger heterotypic interaction only requires one RRM to be activated, therefore, having an early-activated RRM is highly favorable for condensation, which will explain that RRM123 condenses earlier than RRM222.

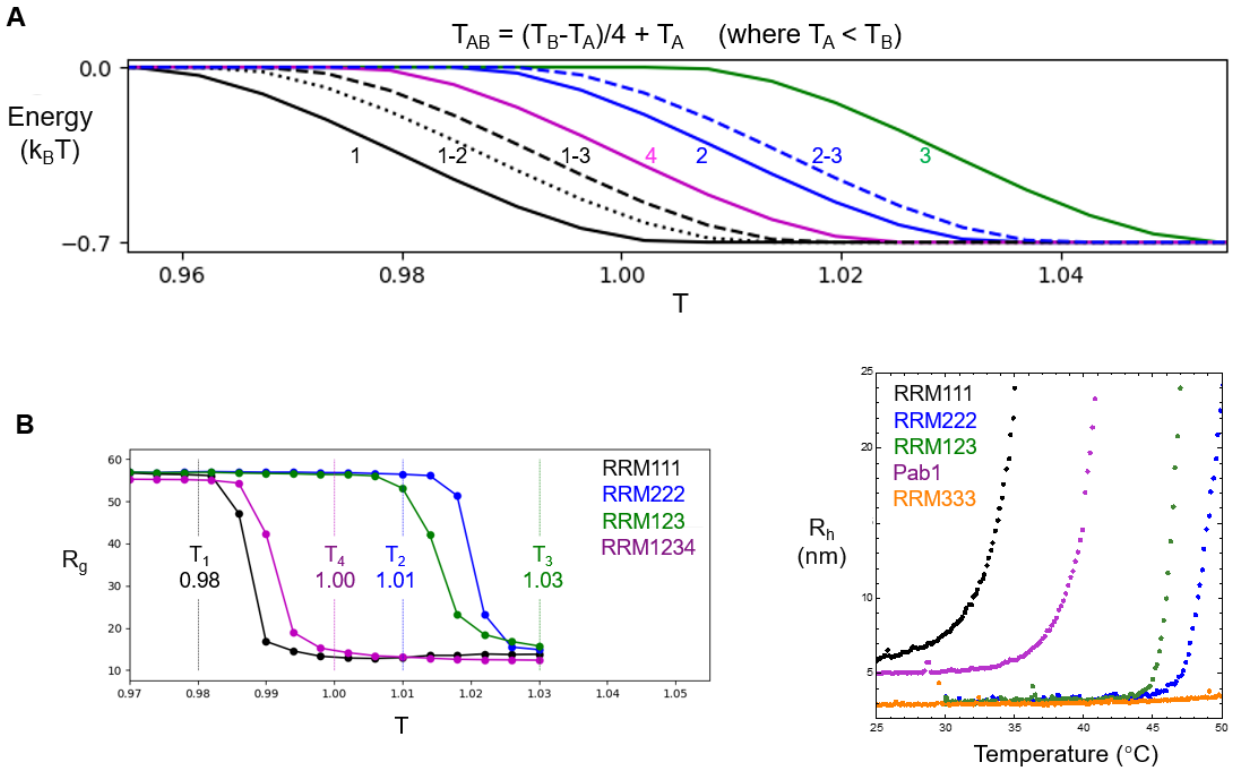


Figure 4.19: Scenario 3 with active heterotypic interactions when one RRM is activated first. (A) the interaction between 2 RRMs goes to the activated energy level at a quarter-way point between the temperatures. (B) Simulation output (left) and experimental DLS data (right).

Next we simulated the co-demixing experiments. We tried 2 groups based on experiments: RRM123 + RRM1234, and RRM222 + RRM1234, and calculated $\%monomer$ rather than R_g as the output. $\%monomer$ is the percentage of non-condensing molecules, therefore a higher $\%monomer$ indicates a lower level of co-demixing.

The simulation recapitulated the following experimental phenomena (Figure 4.20):

1. When mixed with Pab1, both RRM123 and RRM222 have co-demixing before their own T_{demix}: $\%monomer$ of RRM123 started to decrease at lower temperature when mixed with Pab1 (cyan) compared to when RRM123 is by itself without Pab1 (green). Similarly, $\%monomer$ of RRM222 when mixed with Pab1 (red) goes down at lower temperature than that of RRM222 by itself (blue).

2. RRM123 (cyan) co-demixes at lower temperature than RRM222 (red), but at higher

temperature $T > T_2$, co-demixing of RRM222 with Pab1 is enhanced and exceeds co-demixing of RRM123 with Pab1.

3. RRM222 inhibits Pab1's condensation (red dashed, compared to purple) at a lower temperature, but the inhibition is gone after RRM2 becomes activated at $T > T_2$.

It is observed in the simulation but not in the experiments that RRM123 is inhibitive to Pab1's condensation at lower temperatures ($T_1 < T \approx T_4 < T_2$). A plausible explanation is that our experiment temperatures (42 and 46 °C) are out of this simulated temperature range where the inhibition can be seen.

While the model we used is greatly simplified and does not include more complicated cases (e.g. different pairs of heterotypic interactions can be of different energy levels; some RRM interactions might be inhibitive; linkers may form interactions), it provides useful information that RRM interactions are likely non-specific. The simulation also suggests that heterotypic interactions can occur below the higher of the two RRMs' activation temperatures, as in this scenario it yields results qualitatively consistent with results from DLS and co-demixing experiments, with the only input information that 4 RRMs have different activation temperatures $T_1 < T_4 < T_2 < T_3$, which possibly results from that the activation of the first activable RRM may facilitate the activation of the second RRM. Another possible scenario is that there is a gradient of the interaction strength between two RRMs depending on the level of unfolding. An activated RRM1 can have weaker interaction with a not fully activated RRM2 (or other RRM type) which has some low level partial-unfolding.

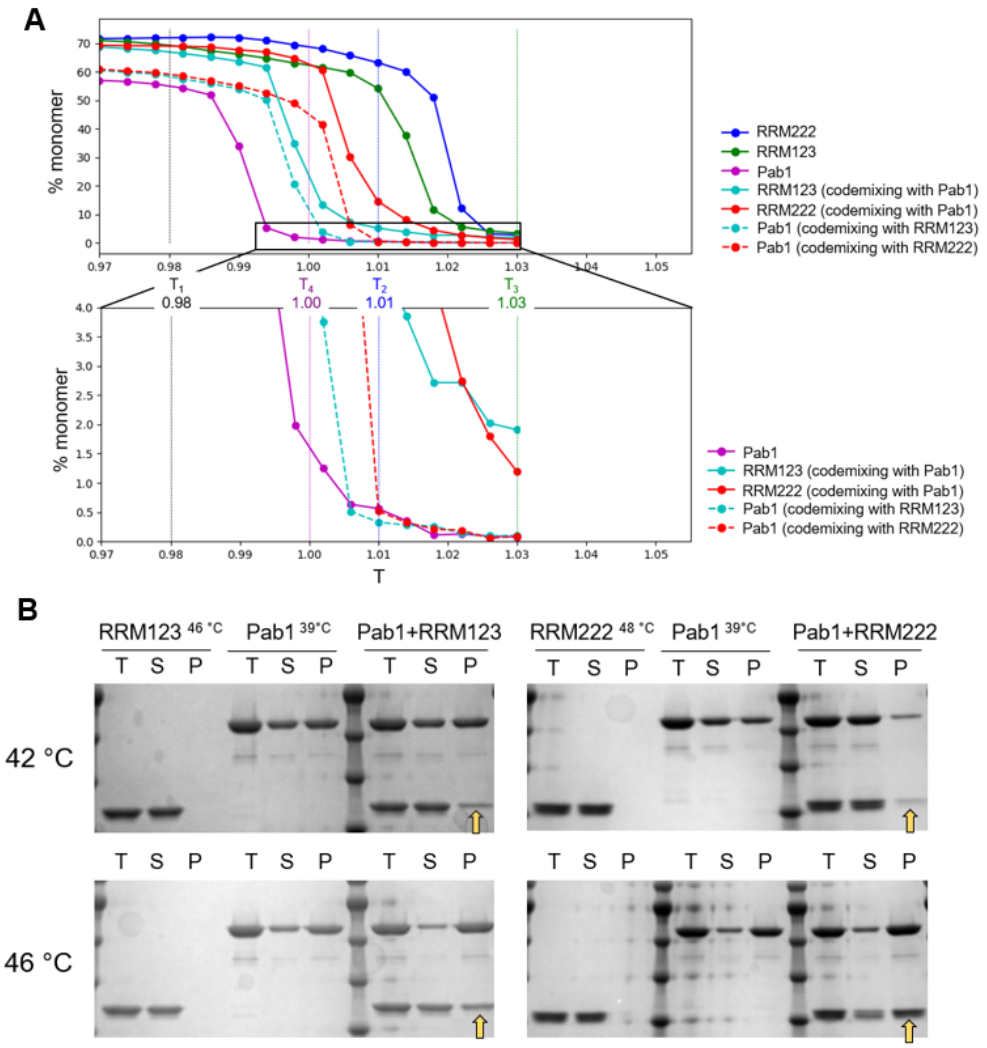


Figure 4.20: Co-demixing of RRM123 and RRM222 with Pab1 (A) Simulation under scenario 3 (B) Experimental data.

4.2 Methods

FRET

Pab1 or cysteine mutants were exchanged into buffer containing 20 mM HEPES, pH 7.0, 150 mM KCl and 1 mM TCEP for reduction. Alexa488 and/or 594 C5 maleimide (Thermo Fisher Scientific) was resuspended in DMSO to 5 mg/ml and added to the protein to a final dye:protein ratio of 10:1. The reaction was protected from light and rotated on a

rotisserie overnight at 4 °C. Unreacted dye was removed by extensive dialysis. Labeling efficiency was between 40%-65% determined by absorbance where dyes' absorbances at 280 nm were corrected. Samples were injected into a quartz cuvette with cap and measured with a Horiba Fluorolog-3 machine with Synapse OE-CCD Array Detector and a temperature control module. Excitation wavelength was set at 478 nm. For each different temperature measurement, sample was equilibrated for 15 min.

SAXS

Data were collected at the BioCAT beamline at the Advanced Photon Source (Argonne National Lab) using in-line SEC-SAXS measurement. A concentrated protein sample (∼3 mg/mL) was injected onto SEC system and flew through a GE Lifesciences Superdex-200 size exclusion column. Data was collected and inspected using RAW software where monomeric peak was selected for which scattering data was integrated and exported. Giunier fit (R_g) and IFT (P(r)) calculation is done in GNOM software.

CHAPTER 5

CONCLUSIONS, OPEN QUESTIONS, AND FUTURE DIRECTIONS

Our studies have elucidated the mechanism of Pab1's condensation. Multiple regions of RRMs are destabilized or unfolded upon condensation whereas linkers, P domain and some regions of RRMs can become more stable. We proposed a "sequential activation" model where RRMs become activated, partially unfolding, and participate in condensation at different temperatures. Here I address some remaining open questions and potential working directions towards those questions.

5.0.1 Interactions between RRMs

The facts that RRM222 and all single-domain-deletion mutant still condense imply that no domain is essential for condensation [57]. Therefore, there should be no specific interaction strictly required for condensation. While there is also evidence from co-demixing data that RRMs can form non-specific heterotypic pairs, it is unclear why RRM222 inhibits RRM113's condensation. Are heterotypic interactions different from homotypic interactions in terms of contribution to condensation? Do heterotypic or homotypic interactions form stochastically or there will be a preference and/or an optimization of interactions in the condensates?

XL-MS

As we have shown that Pab1 condensates are amenable to mass spectrometry based techniques, resolving the details of RRM interactions, potentially by XL-MS (Figure 5.1), would be a reasonable step. However, the data may be highly heterogeneous which will impede the analysis.

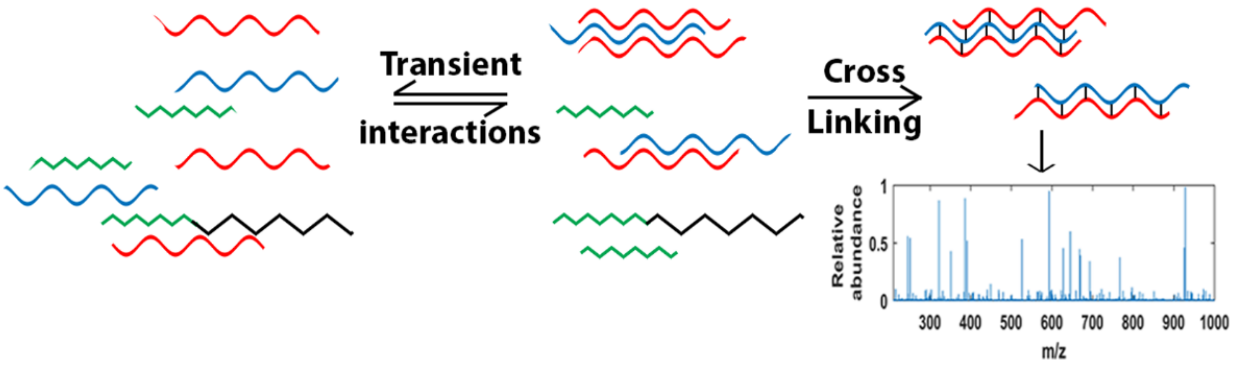


Figure 5.1: Schematic of XL-MS. Transient interaction species are cross-linked and read on mass spectrometer. Figure is from [26].

High throughput co-demixing

As shown in previous chapters, co-demixing can provide information on RRM interactions as one can make constructs of different compositions. Designing high throughput co-demixing experiments to extend to a broader matrix of construct pairs will greatly facilitate the understanding of RRM interactions. A possible experimental set-up is to attach a fluorescence tag to protein, subject samples to condensation conditions followed by a sucrose gradient fractionation. Co-demixing levels can be measured on plate reader as co-fluorescence.

5.0.2 Study the role of P domain by X-ray footprinting/Mass Spectrometry

Pab1's CTD has a peptide-binding motif [39], which unfolds upon condensation. A hypothesis is that the hydrophobic P domain may interact with CTD. From the HDX, P domain is largely exchanging as an unstructured chain in the monomeric state, and in the condensates, some region of P domain has slowed HDX. A limitation of HDX is that it probes solvent accessibility of the backbone amide proton while providing little information about residue side chains. Therefore, for an unstructured IDR P domain, HDX provides less information on how it mediates interactions through the side chain. X-ray footprinting/Mass Spectrometry (XRMS) is a good complementary to HDX as it provides residue-level information on

the solvent accessibility of protein side chains. Another advantage of XRMS is that it can be applied *in vivo* as the covalent modification is not restricted by back exchange as in HDX [1]. Studying Pab1's condensation *in vivo*, and potentially the dispersal of Pab1 condensates by heat shock proteins [69] will greatly aid the understanding of the stress response in a more biology-relevant context.

5.0.3 Is there a directional effect?

That RRM213 condenses similarly to RRM123, but with a 2 °C earlier T_{demix} leads us to conjecture that there is a directional effect involving the arrangement of the RRMs, which influences the optimization of interactions in the condensates and therefore may influence T_{demix}. As linker1 is the shortest linker (9 residues) compared to other linkers, RRM1 and RRM2 are in closer proximity than other RRMs. Does Pab1's condensation require RRM1 and RRM2 to be adjacent to each other? Will RRM132 condense differently from RRM123? Are the first RRM and the last RRM have different behaviors influenced by their neighboring linkers (N-terminal and P domain)? Future studies are needed to investigate this possibility.

5.0.4 Valency and crosslinking in the condensates

The canonical view is that a valency of 2 enables chain formation while a valency above 2 enables the formation of a 3D network. For Pab1, RRM4-P-C construct is non-condensing which implies that only 1 activable RRM is insufficient for condensation. RRM123 is considered to contain two activable domains (RRM3 does not activate until ∼50 °C) and condenses at ∼45 °C. R12 also has a condensation-like behavior. These observations suggest that Pab1 constructs sufficiently condense when there are two activated RRM domains. It is unclear if Pab1 condensates have a chain-like backbone, augmented by other weaker quinary interactions. Whether and how the linkers and RRM3 assist the condensation network (e.g. are they considered to be an extra valency?) remain to be answered.

According to our model, the level of crosslinking should increase with temperature and

lead to an increase in condensate stability and morphology, as the activation of each domain increases the valency. In principle, the condensates would successively change from dimers to chains and then to 3D networks (Figure 5.2). Measuring size distributions of different RRM constructs at different temperatures would be informative for this question. EM would be a useful tool to probe the morphology of different condensates as well.

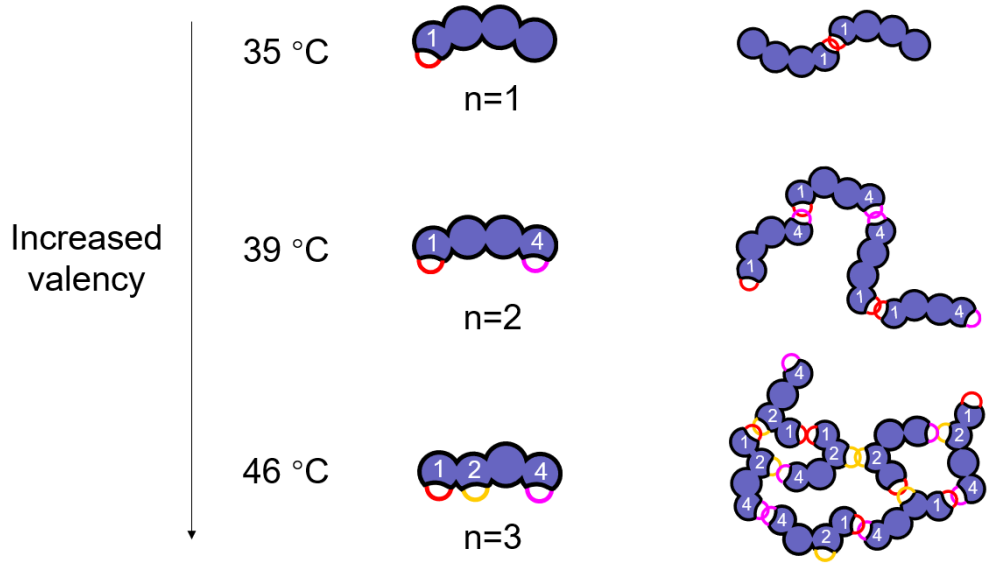


Figure 5.2: Thermodynamic specificity suggests a change in condensate morphology with increasing temperature

5.0.5 Are RRMs' differences evolutionarily conserved?

RRMs in Pab1 have different RNA-binding affinities with RRM1 has the highest affinity [13]. In the condensates, RRMs also show different levels of unfolding and participation in condensation (Figure 2.10, 2.11, 2.12), controlled by their thermodynamic properties. Moreover, RRMs have different unfolding patterns upon condensation (Figure 5.3A), with RRM1 and 4 resembling each other whereas RRM2 has a significantly different unfolding/stabilization pattern on the helices and hairpins. Moreover, RRMs have similar overall compositions of amino acids, but the distribution can be different locally for secondary structure elements

(e.g. the helices, Figure 5.3B).

An open question here is whether the RRMs' differences are evolutionarily conserved? Is there any evolutionary benefit for Pab1 to have 4 different RRMs instead of 4 copies of the same RRM? Does Pab1 become a "differential temperature sensor" to obtain a broader tunable range of temperature sensitivity and therefore convey higher adaptivity to thermally variable environments? For example, psychrophilic PABP has the same RRM1 and RRM2 as Pab1. Mostly by mutating its RRM3, additionally with a single mutation on RRM4 and a few on terminal tails, its T_{demix} is lowered by 7 °C which is adaptive to the low temperature environment. It could be beneficial to be able to tune the condensation by mutating fewer RRMs to preserve other RRM functions such as RNA-binding.

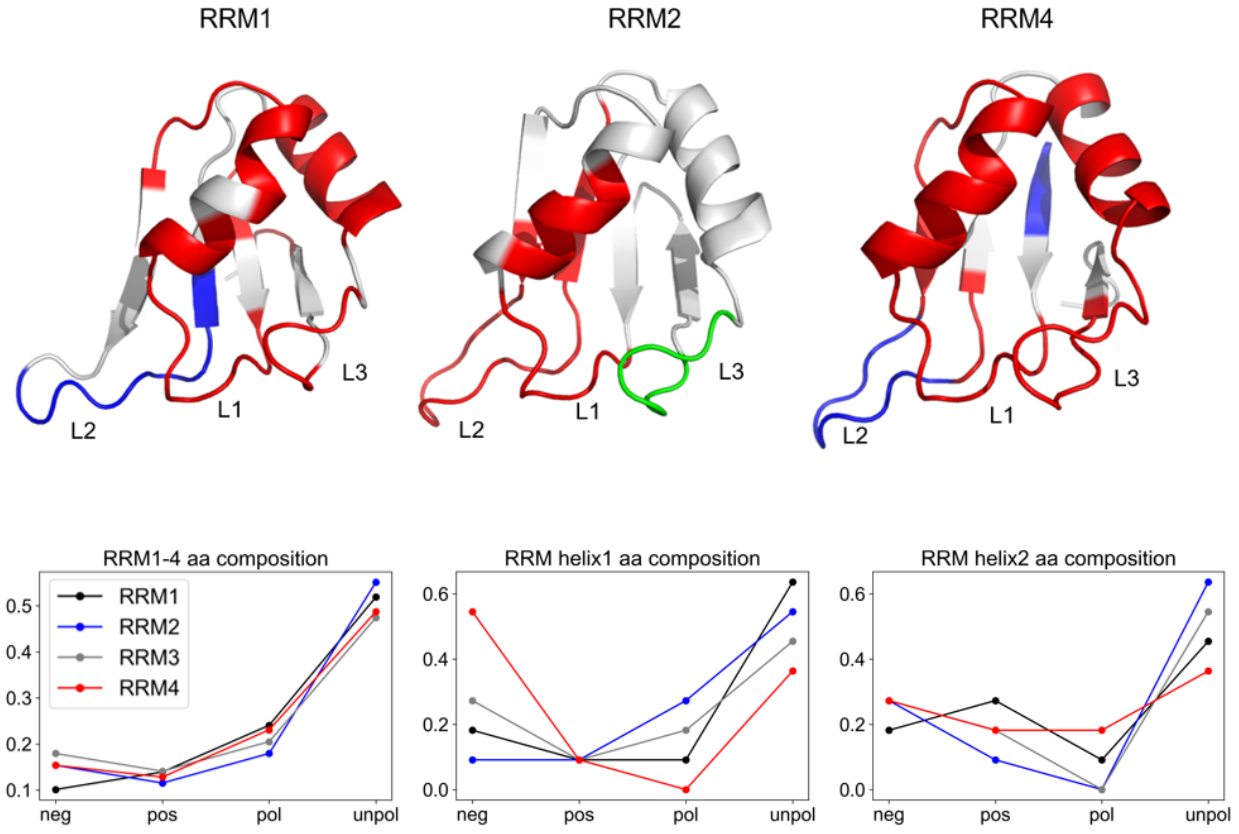


Figure 5.3: Comparison between RRMs' HDX behavior and amino acid compositions. (A) HDX mapped onto RRM structure. Red regions are unfolded while blue is stabilized. Green region has similar HDX rate in the condensates compared to monomeric state. (B) Residue composition of 4 RRMs (overall, helix1, helix2).

5.0.6 Conclusions

Together, this thesis uncovers the molecular mechanism of the stress-triggered condensation of Pab1. We identified the molecular factors of Pab1's condensation by a collection of biophysical measurements. Particularly, we successfully applied HDX-MS to Pab1 condensates and showed that HDX-MS fills the technological gap as a high-resolution analysis tool for biomolecular condensates. We found that RRMs are the key to condensation and they partially unfold above different temperature thresholds. We proposed a "sequencial activation" model for Pab1's condensation and term the mechanism "thermodynamic specificity" wherein an activated RRM is likely to strongly interact only with other activated RRMs rather than inactivated ones.

Our data uncovers that partial unfolding underlies Pab1's condensation process, which supports the view that disorder and quinary interactions underlie the condensation of Pab1 as they do in other systems that condense, just that partial unfolding is an early step of the process. The feature distinguishing Pab1's condensation is the temperature dependent domain activation which results in a more complex behavior as compared to typical biomolecular condensation which can be well described using mean field approximations and Flory-Huggins theory [22, 28].

Pab1's autonomous condensation upon mild stress and its role in translational control [38], together with other stress-induced proteins, including chaperones which effectively disperse condensed Pab1 [69], synthesize a unique and effective cellular stress response. Our findings elucidate how Pab1's structure endows it to be a central stress sensor - Pab1 responds to both pH and thermal stresses with special sensitivity contributed by the "thermodynamic specificity" of RRMs.

REFERENCES

[1] Tadepalli Adilakshmi, Richard A Lease, and Sarah A Woodson. Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. *Nucleic acids research*, 34(8):e64–e64, 2006.

[2] Paul Anderson and Nancy Kedersha. RNA granules. *Journal of Cell Biology*, 172(6):803–808, March 2006.

[3] Daniel Axelrod, DE Koppel, J Schlessinger, Elliot Elson, and Watt W Webb. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical journal*, 16(9):1055–1069, 1976.

[4] Yawen Bai, John S Milne, Leland Mayne, and S Walter Englander. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics*, 17(1):75–86, 1993.

[5] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18(5):285–298, May 2017.

[6] Salman F Banani, Allyson M Rice, William B Peeples, Yuan Lin, Saumya Jain, Roy Parker, and Michael K Rosen. Compositional control of phase-separated cellular bodies. *Cell*, 166(3):651–663, 2016.

[7] Clifford P. Brangwynne, Christian R. Eckmann, David S. Courson, Agata Rybarska, Carsten Hoege, Jbin Gharakhani, Frank Jlicher, and Anthony A. Hyman. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, 324(5935):1729–1732, June 2009. Publisher: American Association for the Advancement of Science.

[8] Clifford P. Brangwynne, Timothy J. Mitchison, and Anthony A. Hyman. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proceedings of the National Academy of Sciences*, 108(11):4334–4339, March 2011.

[9] John Cavanagh, Wayne J Fairbrother, Arthur G Palmer III, and Nicholas J Skelton. *Protein NMR spectroscopy: principles and practice*. Academic press, 1996.

[10] Valeria Cherkasov, Sarah Hofmann, Silke Druffel-Augustin, Axel Mogk, Jens Tyedmers, Georg Stoecklin, and Bernd Bukau. Coordination of translational control and protein homeostasis during severe heat stress. *Current biology*, 23(24):2452–2462, 2013.

[11] Jeong-Mo Choi, Alex S. Holehouse, and Rohit V. Pappu. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics*, 49(1):107–133, May 2020.

[12] Julie A Deardorff and Alan B Sachs. Differential effects of aromatic and charged residue substitutions in the rna binding domains of the yeast poly (a)-binding protein. *Journal of molecular biology*, 269(1):67–81, 1997.

[13] Rahul C Deo, Jeffrey B Bonanno, Nahum Sonenberg, and Stephen K Burley. Recognition of polyadenylate rna by the poly (a)-binding protein. *Cell*, 98(6):835–845, 1999.

[14] A. Keith Dunker, J. David Lawson, Celeste J Brown, Ryan M Williams, Pedro Romero, Jeong S Oh, Christopher J Oldfield, Andrew M Campen, Catherine M Ratliff, Kerry W Hipps, Juan Ausio, Mark S Nissen, Raymond Reeves, ChulHee Kang, Charles R Kissinger, Robert W Bailey, Michael D Griswold, Wah Chiu, Ethan C Garner, and Zoran Obradovic. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1):26–59, February 2001.

[15] H. Jane Dyson and Peter E. Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3):197–208, March 2005.

[16] SJ Edelstein. Patterns in the quinary structures of proteins. plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. *Biophysical Journal*, 32(1):347–360, 1980.

[17] Shana Elbaum-Garfinkle, Younghoon Kim, Krzysztof Szczepaniak, Carlos Chih-Hsiung Chen, Christian R. Eckmann, Sua Myong, and Clifford P. Brangwynne. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences*, 112(23):7189–7194, June 2015.

[18] Avigdor Eldar and Michael B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010.

[19] R John Ellis. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current opinion in structural biology*, 11(1):114–119, 2001.

[20] S. W. Englander, L. Mayne, Y. Bai, and T. R. Sosnick. Hydrogen exchange: the modern legacy of Linderstrm-Lang. *Protein Science : A Publication of the Protein Society*, 6(5):1101–1109, May 1997.

[21] S Walter Englander, Tobin R Sosnick, Joan J Englander, and Leland Mayne. Mechanisms and uses of hydrogen exchange. *Current opinion in structural biology*, 6(1):18–23, 1996.

[22] Paul J. Flory. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10(1):51–61, January 1942. Publisher: American Institute of Physics.

[23] Tobias M Franks and Jens Lykke-Andersen. The control of mrna decapping and p-body formation. *Molecular cell*, 32(5):605–615, 2008.

[24] Titus M. Franzmann, Marcus Jahnel, Andrei Pozniakovsky, Julia Mahamid, Alex S. Holehouse, Elisabeth Nske, Doris Richter, Wolfgang Baumeister, Stephan W. Grill,

Rohit V. Pappu, Anthony A. Hyman, and Simon Alberti. Phase separation of a yeast prion protein promotes cellular fitness. *Science*, 359(6371):eaao5654, January 2018.

[25] Ho Yee Joyce Fung, Melissa Birol, and Elizabeth Rhoades. Idps in macromolecular complexes: the roles of multivalent interactions in diverse assemblies. *Current opinion in structural biology*, 49:36–43, 2018.

[26] Drishti Guin and Martin Gruebele. Weak Chemical Interactions That Drive Protein Evolution: Crowding, Sticking, and Quinary Structure in Folding and Function. *Chemical Reviews*, 119(18):10691–10717, September 2019.

[27] Tyler S Harmon, Alex S Holehouse, Michael K Rosen, and Rohit V Pappu. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife*, 6:e30294, November 2017. Publisher: eLife Sciences Publications, Ltd.

[28] Maurice L. Huggins. Some Properties of Solutions of Long-chain Compounds. *The Journal of Physical Chemistry*, 46(1):151–158, January 1942. Publisher: American Chemical Society.

[29] Tsang-Lin Hwang, Susumu Mori, A. J. Shaka, and Peter C. M. van Zijl. Application of Phase-Modulated CLEAN Chemical EXchange Spectroscopy (CLEANEX-PM) to Detect WaterProtein Proton Exchange and Intermolecular NOEs. *Journal of the American Chemical Society*, 119(26):6203–6204, July 1997. Publisher: American Chemical Society.

[30] Saumya Jain, Joshua R Wheeler, Robert W Walters, Anurag Agrawal, Anthony Barsic, and Roy Parker. Atpase-modulated stress granules contain a diverse proteome and substructure. *Cell*, 164(3):487–498, 2016.

[31] Ellie I. James, Taylor A. Murphree, Clint Vorauer, John R. Engen, and Miklos Guttman.

Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems. *Chemical Reviews*, 122(8):7562–7623, April 2022.

[32] John M Jumper, Nabil F Faruk, Karl F Freed, and Tobin R Sosnick. Trajectory-based training enables protein simulations with accurate folding and boltzmann ensembles in cpu-hours. *PLoS computational biology*, 14(12):e1006578, 2018.

[33] Masato Kato, Tina W Han, Shanhai Xie, Kevin Shi, Xinlin Du, Leeju C Wu, Hamid Mirzaei, Elizabeth J Goldsmith, Jamie Longgood, Jimin Pei, et al. Cell-free formation of rna granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*, 149(4):753–767, 2012.

[34] Nancy Kedersha and Paul Anderson. Stress granules: sites of mrna triage that regulate mrna stability and translatability. *Biochemical Society Transactions*, 30(6):963–969, 2002.

[35] Nancy Kedersha, Michael R Cho, Wei Li, Patrick W Yacono, Samantha Chen, Natalie Gilks, David E Golan, and Paul Anderson. Dynamic shuttling of tia-1 accompanies the recruitment of mrna to mammalian stress granules. *The Journal of cell biology*, 151(6):1257–1268, 2000.

[36] Nancy L Kedersha, Mita Gupta, Wei Li, Ira Miller, and Paul Anderson. Rna-binding proteins tia-1 and tiar link the phosphorylation of eif-2α to the assembly of mammalian stress granules. *The Journal of cell biology*, 147(7):1431–1442, 1999.

[37] Indu Kheterpal, Shaolian Zhou, Kelsey D. Cook, and Ronald Wetzel. A amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proceedings of the National Academy of Sciences*, 97(25):13597–13601, December 2000. Publisher: Proceedings of the National Academy of Sciences.

[38] Hemant K Kini, Ian M Silverman, Xinjun Ji, Brian D Gregory, and Stephen A Lieb-

haber. Cytoplasmic poly (a) binding protein-1 binds to genomically encoded sequences within mammalian mrnas. *Rna*, 22(1):61–74, 2016.

[39] Guennadi Kozlov, Nadeem Siddiqui, Stephane Coillet-Matillon, Jean-Franois Trempe, Irena Ekiel, Tara Sprules, and Kalle Gehring. Solution Structure of the Orphan PABC Domain fromSaccharomyces cerevisiae Poly(A)-binding Protein. *Journal of Biological Chemistry*, 277(25):22822–22828, June 2002.

[40] Sonja Kroschwald, Shovamayee Maharana, Daniel Mateju, Liliana Malinovska, Elisabeth Nüske, Ina Poser, Doris Richter, and Simon Alberti. Promiscuous interactions and protein disaggregases determine the material state of stress-inducible rnp granules. *elife*, 4:e06807, 2015.

[41] Mavis Jiarong Li, Miklos Guttman, and William M Atkins. Conformational dynamics of p-glycoprotein in lipid nanodiscs and detergent micelles reveal complex motions on a wide time scale. *Journal of Biological Chemistry*, 293(17):6297–6307, 2018.

[42] Pilong Li, Sudeep Banjade, Hui-Chun Cheng, Soyeon Kim, Baoyu Chen, Liang Guo, Marc Llaguno, Javoris V Hollingsworth, David S King, Salman F Banani, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483(7389):336–340, 2012.

[43] Yuan Lin, DavidS.W. Protter, MichaelK. Rosen, and Roy Parker. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell*, 60(2):208–219, October 2015.

[44] Jidong Liu, Fabiola V Rivas, James Wohlschlegel, John R Yates, Roy Parker, and Gregory J Hannon. A role for the p-body component gw182 in microrna function. *Nature cell biology*, 7(12):1261–1266, 2005.

[45] David A Mangus, Matthew C Evans, and Allan Jacobson. Poly (a)-binding proteins:

multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome biology*, 4(7):1–14, 2003.

[46] Christophe Maris, Cyril Dominguez, and Frdric H.-T. Allain. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression: The RRM domain, a plastic RNA-binding platform. *FEBS Journal*, 272(9):2118–2131, April 2005.

[47] Erik W. Martin and Tanja Mittag. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry*, 57(17):2478–2487, May 2018.

[48] Leland Mayne, Zhong-Yuan Kan, Palaniappan Sevugan Chetty, Alec Ricciuti, Benjamin T. Walters, and S. Walter Englander. Many Overlapping Peptides for Protein Hydrogen Exchange Experiments by the Fragment Separation-Mass Spectrometry Method. *Journal of the American Society for Mass Spectrometry*, 22(11):s13361–011–0235–4, November 2011.

[49] Edward McEwen, Nancy Kedersha, Benbo Song, Donalyn Scheuner, Natalie Gilks, Anping Han, Jane-Jane Chen, Paul Anderson, and Randal J Kaufman. Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. *Journal of Biological Chemistry*, 280(17):16925–16933, 2005.

[50] Patrick S Merkle, Kamil Gotfryd, Michel A Cuendet, Katrine Z Leth-Espensen, Ulrik Gether, Claus J Loland, and Kasper D Rand. Substrate-modulated unwinding of transmembrane helices in the nss transporter leut. *Science advances*, 4(5):eaar6179, 2018.

[51] Amandine Molliex, Jamshid Temirov, Jihun Lee, Maura Coughlin, Anderson P Kanagaraj, Hong Joo Kim, Tanja Mittag, and J Paul Taylor. Phase separation by low com-

plexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*, 163(1):123–133, 2015.

[52] Christopher Frederick Mugler, Maria Hondele, Stephanie Heinrich, Ruchika Sachdev, Pascal Vallotton, Adriana Y Koek, Leon Y Chan, and Karsten Weis. Atpase activity of the dead-box protein dhh1 controls processing body formation. *Elife*, 5:e18746, 2016.

[53] Timothy J Nott, Evangelia Petsalaki, Patrick Farber, Dylan Jervis, Eden Fussner, Anne Plochowietz, Timothy D Craggs, David P Bazett-Jones, Tony Pawson, Julie D Forman-Kay, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell*, 57(5):936–947, 2015.

[54] Chi W Pak, Martyna Kosno, Alex S Holehouse, Shae B Padrick, Anuradha Mittal, Rustam Ali, Ali A Yunus, David R Liu, Rohit V Pappu, and Michael K Rosen. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Molecular cell*, 63(1):72–85, 2016.

[55] Avinash Patel, HyunO. Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, MarcoY. Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, TitusM. Franzmann, Andrej Pozniakovski, Ina Poser, Nicola Maghelli, LoicA. Royer, Martin Weigert, EugeneW. Myers, Stephan Grill, David Drechsel, AnthonyA. Hyman, and Simon Alberti. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, 162(5):1066–1077, August 2015.

[56] Robert D Phair and Tom Misteli. High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778):604–609, 2000.

[57] Joshua A. Riback, Christopher D. Katanski, Jamie L. Kear-Scott, Evgeny V. Pilipenko, Alexandra E. Rojek, Tobin R. Sosnick, and D. Allan Drummond. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*, 168(6):1028–1040.e19, March 2017.

[58] Shambaditya Saha, Christoph A Weber, Marco Nousch, Omar Adame-Arana, Carsten Hoege, Marco Y Hein, Erin Osborne-Nishimura, Julia Mahamid, Marcus Jahnel, Louise Jawerth, et al. Polar positioning of phase-separated liquid compartments in cells regulated by an mrna competition mechanism. *Cell*, 166(6):1572–1584, 2016.

[59] Ingmar B Schäfer, Masami Yamashita, Jan Michael Schuller, Steffen Schüssler, Peter Reichelt, Mike Strauss, and Elena Conti. Molecular basis for poly (a) rnp architecture and recognition by the pan2-pan3 deadenylase. *Cell*, 177(6):1619–1631, 2019.

[60] Yongdae Shin and Clifford P. Brangwynne. Liquid phase condensation in cell physiology and disease. *Science*, 357(6357):eaaf4382, September 2017.

[61] Jarrett Smith, Deepika Calidas, Helen Schmidt, Tu Lu, Dominique Rasoloson, and Geraldine Seydoux. Spatial patterning of p granules by rna-induced phase separation of the intrinsically-disordered protein meg-3. *Elife*, 5:e21337, 2016.

[62] Chun So, Shiya Cheng, and Melina Schuh. Phase separation during germline development. *Trends in cell biology*, 31(4):254–268, 2021.

[63] Sylvie Souquere, Stéphanie Mollet, Michel Kress, François Dautry, Gérard Pierron, and Dominique Weil. Unravelling the ultrastructure of stress granules and associated p-bodies in human cells. *Journal of cell science*, 122(20):3619–3626, 2009.

[64] Haofeng Sun, Lingyun Ma, Leyu Wang, Peng Xiao, Hongmei Li, Min Zhou, and Dewei Song. Research advances in hydrogen–deuterium exchange mass spectrometry for protein epitope mapping. *Analytical and Bioanalytical Chemistry*, 413(9):2345–2359, 2021.

[65] Jacob Verghese, Jennifer Abrams, Yanyu Wang, and Kevin A Morano. Biology of the heat shock response and protein chaperones: budding yeast (saccharomyces cerevisiae) as a model system. *Microbiology and Molecular Biology Reviews*, 76(2):115–158, 2012.

[66] Edward WJ Wallace, Jamie L Kear-Scott, Evgeny V Pilipenko, Michael H Schwartz, Pawel R Laskowski, Alexandra E Rojek, Christopher D Katanski, Joshua A Riback, Michael F Dion, Alexander M Franks, et al. Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. *Cell*, 162(6):1286–1298, 2015.

[67] Jie Wang, Jeong-Mo Choi, Alex S. Holehouse, Hyun O. Lee, Xiaojie Zhang, Marcus Jahnel, Shovamayee Maharana, Rgis Lemaitre, Andrei Pozniakovsky, David Drechsel, Ina Poser, Rohit V. Pappu, Simon Alberti, and Anthony A. Hyman. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, 174(3):688–699.e16, July 2018.

[68] Peiguo Yang, Cécile Mathieu, Regina-Maria Kolaitis, Peipei Zhang, James Messing, Ugur Yurtsever, Zemin Yang, Jinjun Wu, Yuxin Li, Qingfei Pan, et al. G3bp1 is a tunable switch that triggers phase separation to assemble stress granules. *Cell*, 181(2):325–345, 2020.

[69] Haneul Yoo, Jared A. M. Bard, Evgeny V. Pilipenko, and D. Allan Drummond. Chaperones directly and efficiently disperse stress-triggered biomolecular condensates. *Molecular Cell*, 82(4):741–755.e11, February 2022.

[70] Zhongqi Zhang and David L. Smith. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation: Amide hydrogen exchange by mass spectrometry. *Protein Science*, 2(4):522–531, April 1993.