Winter 12-15-2022

The Physics Of Associative Polymers And Applications To Biomolecular Condensates

Furqan Dar
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Physics Commons

Recommended Citation
Dar, Furqan, "The Physics Of Associative Polymers And Applications To Biomolecular Condensates" (2022). Arts & Sciences Electronic Theses and Dissertations. 2770.
https://openscholarship.wustl.edu/art_sci_etds/2770

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
WASHINGTON UNIVERSITY IN ST. LOUIS

College of Arts & Sciences
Department of Physics

Dissertation Examination Committee:
Anders E. Carlsson, Chair
Rohit V. Pappu, Co-chair
Alex S. Holehouse
James Mertens
Shankar Mukherji

The Physics Of Associative Polymers
And Applications To Biomolecular Condensates
by
Furqan Dar

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2022
St. Louis, Missouri
# Table of Contents

List of Figures........................................................................................................ viii
List of Tables ........................................................................................................ xiv
Acknowledgments................................................................................................... xv
Abstract .................................................................................................................. xviii

1. Introduction ........................................................................................................ 1
   1.1 Preamble ...................................................................................................... 1
   1.2 Biomolecular Condensates ....................................................................... 2
      1.2.1 Constituents Of Biomolecular Condensates .................................. 5
   1.3 Phase Separation Coupled To Percolation: Theoretical Background .......... 7
   1.4 Numerically Investigating PSCP .............................................................. 12
      1.4.1 Field Theoretic Approaches ....................................................... 12
      1.4.2 Explicitly Modeling Macromolecules ........................................... 14
   1.5 Scope Of The Thesis ............................................................................... 17
   1.6 References ............................................................................................... 20

2. LaSSI: Lattice Simulation Engine For Sticker And Spacer Interactions ... 31
   2.1 Preamble ................................................................................................. 31
   2.2 Introduction ............................................................................................. 32
   2.3 LaSSI Model ........................................................................................... 39
      2.3.1 Extension To Bond Fluctuation Models ....................................... 40
   2.4 Move Set .................................................................................................. 42
      2.4.1 Biased Monte Carlo Sampling ..................................................... 43
      2.4.2 Sticker Rotation Moves ............................................................... 45
      2.4.3 Local Moves .............................................................................. 47
# 2.4.4 Co-Local Moves ................................................................. 50
# 2.4.5 Slithering Snake Moves ................................................... 51
# 2.4.6 Multi Local Moves ........................................................... 53
# 2.4.7 Translation Moves ............................................................ 55
# 2.4.8 Old Cluster Translation Moves ........................................... 56
# 2.4.9 New Cluster Translation Moves .......................................... 57
# 2.4.10 Branched Rotation Moves ................................................. 59
# 2.4.11 Pivot Moves .................................................................... 60
# 2.4.12 Double Pivot Moves ........................................................ 62

2.5 References ............................................................................... 65

## 3. Using LaSSI To Investigate The Phase Behavior Of Multivalent Proteins 77

3.1 Preamble ............................................................................... 77
3.2 Introduction ............................................................................ 77
  3.2.1 Building Intuition For Setting Move Frequencies ....................... 79
3.3 Linear Multivalent Proteins ....................................................... 81
  3.3.1 FUS As An Archetypal Linear Multivalent Protein ....................... 81
  3.3.2 Detecting Phase Transitions Using Distribution Functions .......... 83
  3.3.3 Quantitative Assessment Of Finite Size Effects ......................... 86
  3.3.4 Estimating The Percolation Transition Boundary ....................... 91
  3.3.5 Results And Discussion ..................................................... 93
3.4 Branched Multivalent Proteins ............................................... 99
  3.4.1 N130 & rpL5 As An Archetypal System ..................................... 99
  3.4.2 Results And Discussion ..................................................... 101
3.5 Conclusions ............................................................................ 110
3.6 References .............................................................................. 112

## 4. Ligand Effects On The Phase Separation of Multivalent Macromolecules 119

4.1 Preamble ............................................................................... 119
4.2 Introduction ............................................................................ 120
4.3 Materials And Methods ........................................................... 126
  4.3.1 Coarse-Grained Model ....................................................... 126
4.3.2 Simulation Details ................................................................. 127
4.3.3 Constructing Binodals .......................................................... 129
4.3.4 Generating Site Molecularity Profiles ..................................... 130
4.3.5 Constructing Binding Polynomials ......................................... 131
4.3.6 Calculating PCs From Polyphasic Linkage Theory ................. 133
4.4 Results ..................................................................................... 137
4.4.1 Overall Effects Of Different Types Of Ligands ...................... 137
4.4.2 Effects Of Ligands On Dilute Phase Concentrations .............. 142
4.4.3 Effects Of Ligands On Dense Phase Concentrations ............. 146
4.4.4 Effects Of Ligands On Condensate Structures .................... 149
4.4.5 PCs Of Ligands Alone Cannot Predict Which Phase Ligands Preferentially Bind To ................................................................. 152
4.4.6 Polyphasic Linkage Theory Establishes That PCs Of Ligands Combine Contributions Of Preferential Binding And Local Concentration Effects 155
4.4.7 Effects Of Ligands On Two-Component Systems With Heterotypic Interactions Only ................................................................. 158
4.4.8 Effects Of Ligands On Two-Component Systems With Homotypic And Heterotypic Interactions ....................................................... 163
4.5 Discussion .............................................................................. 166
4.6 References ............................................................................. 170

5. Structure-Function Relationships In Mitochondrial Transcriptional Condensates ........................................................................ 177
5.1 Preamble .................................................................................. 177
5.2 Introduction ........................................................................... 178
5.2.1 The mt-Nucleoid As A Model Functional Condensate .......... 179
5.3 Materials & Methods ................................................................. 180
5.3.1 Cell Culture ......................................................................... 180
5.3.2 siRNA ................................................................................ 181
5.3.3 Western Blot ....................................................................... 181
5.3.4 mtDNA Depletion ............................................................. 182
5.3.5 Fixation And Labelling ....................................................... 182
5.3.6 DNA Template For Transcription Assays ......................... 183
5.3.7 Protein Purification ................................................................. 183
5.3.8 Protein Labelling .................................................................... 184
5.3.9 Equilibrium Phase Separation Assays .......................................... 184
5.3.10 In-vitro Transcription Reaction .................................................. 187
5.3.11 Transcription Assays Using Radiomucletides ................................ 188
5.3.12 Time-course For In-vitro Transcription Reactions........................... 188
5.3.13 Exogenous RNA Experiments .................................................... 189
5.3.14 Droplet Fixation And mt-RNA FISH Labelling ......................... 189
5.3.15 FRAP Experiments ................................................................. 190
5.3.16 Light Microscopy .................................................................... 190
5.3.17 Quantitative image analysis ....................................................... 191
5.3.18 Bioinformatics ........................................................................ 191
5.3.19 Computational Modeling .......................................................... 191

5.4 Experimental Results ......................................................................... 197
5.4.1 Individual Components Of The Mitochondrial Transcription Undergo Phase Separation in-vitro .......................................................... 197
5.4.2 Multicomponent Systems Form Heterogeneously Organized And Dynamically Arrested Condensates in-vitro .............................................. 202
5.4.3 Condensates Support Transcription And Dampen Transcription Rates 204
5.4.4 Newly Synthesized RNA Transcripts Shape Condensate Structures.... 210
5.4.5 Organization Of mt-Nucleoids Is Altered Upon Depletion Of Core Transcription Components in-vivo .......................................................... 214

5.5 Computational Results ....................................................................... 220
5.5.1 Matching Experimental Phase Behavior With Coarse-Grained Model. 221
5.5.2 Condensate Reorganization via RNA Transcription ...................... 226
5.5.3 RNA-binding Proteins Can Prevent Reorganization Due To RNA Transcription ........................................................................... 227

5.6 Discussion ....................................................................................... 229
5.7 References ....................................................................................... 234

6. RNA Binding Proteins Form Mesoscale Clusters In Subsaturated Solutions 242
6.1 Preamble ......................................................................................... 242
6.2 Introduction ..................................................................................... 243

6.3 Materials And Methods ........................................................................ 245
6.3.1 Materials ............................................................................... 245
6.3.2 Constructs, Protein Expression And Purification: Method A .......... 247
6.3.3 Constructs, Protein Expression And Purification: Method B .......... 249
6.3.4 Phase Separation Assays ........................................................... 250
6.3.5 Image Collection ..................................................................... 250
6.3.6 Dynamic Light Scattering Experiments (DLS) ......................... 251
6.3.7 Nanoparticle Tracking Analysis (NTA) ................................... 253
6.3.8 Transmission Electron Microscopy (TEM) .............................. 254
6.3.9 Förster Resonance Energy Transfer (FRET) ............................ 254
6.3.10 Anisotropy Measurements ......................................................... 255
6.3.11 Microfluidic Confocal Spectroscopy (MCS) ............................. 256
6.3.12 Computational Modeling .......................................................... 258

6.4 Results ........................................................................................... 260
6.4.1 Macroscopic Phase Separation Is Not Observed In Subsaturated Solutions ................................................................................................................. 260
6.4.2 Clusters Spanning A Range Of Sizes Form In Subsaturated Solutions Of FUS ........................................................................................................ 262
6.4.3 Morphologies Of The Largest Clusters That Form In Subsaturated Solutions ........................................................................................................... 265
6.4.4 Mesoscale Clusters Are Of Low Overall Abundance ............... 266
6.4.5 Clusters Increase In Size With Increasing Concentrations And The Size Distributions Have Heavy Tails ................................................................. 267
6.4.6 Other FET Family Proteins Also Form Clusters In Subsaturated Solutions ........................................................................................................... 269
6.4.7 Clusters Are Reversible And Molecules Exchange Between Clusters .............................. 270
6.4.8 Growth Of Macroscopic Phases Above $c_{\text{sat}}$ ................................. 272
6.4.9 Cluster Formation And Macroscopic Phase Separation Can Be Decoupled ................................................................. 274
6.4.10 Impacts Of Mutagenesis On Cluster Formation And Phase Separation 277
6.4.11 Impact Of The Diversity Of Chemistry Specific Interactions On Cluster Formation And Phase Separation ............................................. 281
6.4.12 Simulations Show That Systems Featuring At Least Two Energy Scales Behave Very Differently From Those With A Single Energy Scale...... 282
6.5 Discussion .......................................................................................... 285
6.5.1 Precedents For Clusters In Subsaturated Solutions ......................... 288
6.5.2 Biological Relevance Of Clusters That Form In Subsaturated Solutions 289
6.6 References .......................................................................................... 291

7. Conclusion & Future Directions ............................................................. 303
7.1 Preamble............................................................................................ 303
7.2 Summary .......................................................................................... 304
7.3 Future Directions.................................................................................. 308
7.4 References.......................................................................................... 315

Appendix A: Supplementary Tables For Chapter 4 ...................................[319]
A.1 References .......................................................................................[327]

Appendix B: Supplementary Figures & Data For Chapter 6 .....................[331]
B.1 Amino Acid Sequences Of Proteins Used.............................................[358]
# List of Figures

| Figure 1.1: | Cartoon Depiction Of Phase Separation Of A Binary System | 4 |
| Figure 1.2: | The Flory-Huggins Entropy Of Mixing For Lattice Polymers | 8 |
| Figure 1.3: | The Flory-Huggins Free Energy Density Of A Binary Solution As A Function Of $\chi$ | 10 |
| Figure 2.1: | Designing A Coarse-Grained Model | 37 |
| Figure 2.2: | The LaSSI model | 41 |
| Figure 2.3: | Sticker Rotation Move | 46 |
| Figure 2.4: | Local Move | 48 |
| Figure 2.5: | Co-Local Move | 51 |
| Figure 2.6: | Slithering Snake Move | 52 |
| Figure 2.7: | Multi-Local Move | 54 |
| Figure 2.8: | Chain Translation Move | 56 |
| Figure 2.9: | Cluster Move Reversibility | 59 |
| Figure 2.10: | Branched Rotation Move | 60 |
| Figure 2.11: | Pivot Move | 61 |
| Figure 2.12: | Double-Pivot Move | 63 |
| Figure 3.1: | Coarse-Grained Representation Of An Archetypal Linear Multivalent System | 82 |
| Figure 3.2: | Using Distributions To Detect Density Transitions | 85 |
| Figure 3.3: | Assessing Finite Size Effects In Simulations With Ideal Chains | 88 |
| Figure 3.4: | Assessing Finite Size Effects In Simulations With Real Chains | 89 |
| Figure 3.5: | Assessing Finite Size Effects For $\tilde{g}(r^*)$ | 91 |
Figure 5.17: Radial Density Profiles As A Function Of Time For All Components Of mt-Nucleoid Transcriptional Machinery (Model A) 226

Figure 5.18: Representative Snapshots Of Simulated Droplets Of Model A As A Function Of Time 227

Figure 5.19: Representative Computational Images Of Simulated Droplets Of Model B As A Function Of Time 228

Figure 5.20: Radial Density Profiles As A Function Of Time For Individual Components Of mt-Nucleoid Transcriptional Machinery (Model B) 229

Figure 6.1: FUS-SNAP Has A Quantifiable \( c_{\text{sat}} \) Whereby Condensates Form Only In Supersaturated Solutions 261

Figure 6.2: DLS Data Show That Clusters Spanning A Range Of Sizes Form In Subsaturated Solutions 263

Figure 6.3: Mesoscale Clusters That Form In Subsaturated Solutions Have Quasi-Spherical Morphologies 265

Figure 6.4: NTA Measurements Show That Mesoscale Clusters Formed In Subsaturated Solutions Have Low Overall Abundance 266

Figure 6.5: Fluorescence Anisotropy And Brightness Based Measurements Show The Presence Of Clusters In Subsaturated Solutions 268

Figure 6.6: Clusters Also Form In Subsaturated Solutions Of FET Family Proteins 269

Figure 6.7: Clusters In Subsaturated Solution Form Via Reversible Associations And Molecules Readily Exchange Between Clusters 271

Figure 6.8: Clusters Grow Into Micron Scale Bodies Above \( c_{\text{sat}} \) 273

Figure 6.9: Autocorrelation Functions From DLS Measurements Collected At The Protein Specific \( c_{\text{sat}} \) Values For hnRNP-A3, EWSR1-SNAP, TAF15-SNAP, AND FUS-SNAP 274

Figure 6.10: Solutes Dissolve Condensates While Having A Minimal Effect On Clusters 276

Figure 6.11: Impact Of pH And Linked Changes To Net Charge On Cluster Formation And Macroscopic Phase Separation Of Untagged FUS 278

Figure 6.12: Mutagenesis Experiments Reveal Chemistry Specific Effects Of Different Residue Types On The Formation Of Mesoscale Clusters In Subsaturated Solutions Of FUS 280
Figure B.14: Control FRET Experiments Performed Using Free Dyes .................[347]

Figure B.15: Increased Concentration Of KCl Weakens Cluster Formation And Macroscopic Phase Separation.................................................................[349]

Figure B.16: Impact Of Mutations Within The PLD And RBD On The Concentration Dependence Of Size Distributions Of Mesoscale Clusters Formed By Untagged FUS.................................................................................[351]

Figure B.17: Impact Of Substituting 24 Arg Residues Within The RBD Of FUS To Gly ..........................................................[351]

Figure B.18: DLS Data Complementing The NTA Distributions Shown In Figure B.17 ..........................................................[353]

Figure B.19: Impact Of Substituting 10 Asp / 4 Glu Residues To Gly Within The RBD Of FUS On The DLS Measured Autocorrelation Functions At Different Concentrations Corresponding To Subsaturated Solutions...[355]

Figure B.20: DLS Data Showing The Chemistry Specific Effects Of Phe To Gly vs. Tyr To Ser Substitutions On The Formation Of Mesoscale Clusters In Subsaturated Solutions Of Untagged FUS........................................[356]

Figure B.21: Comparisons Of Cluster Formation And Macroscopic Phase Separation For The PLD vs. RBD Of FUS ..........................................................[357]

Figure B.22: Comparing Dense Phase Volume Fractions Formed By Homopolymers And Associative Polymers.................................................................[358]
List of Tables

Table 3.1: Simulation parameters for system descriptions. ................................. 78
Table 3.2: Move Frequencies According To Different Types For Linear And Branched Systems................................................................. 80
Table 4.1: MC move frequencies used in simulations........................................ 129
Table 5.1: Compositions for mt-Nucleoid components for binary and ternary mixtures. .................................................................................. 186
Table 5.2: Monte-Carlo Move Frequencies For Coarse-Grained Model Of mt-Nucleoid Condensates.............................................................. 192
Table 5.3: Molecule Numbers For Coarse-Grained Simulations Of Binary Mixtures Of mt-Nucleoid Components.......................................... 194
Table 5.4: Molecule Numbers For Coarse-Grained Simulations Of Full Set Of mt-Nucleoid Components ......................................................... 196
Table 6.1: List of reagents, sources and vendor identifiers if any. ....................... 245
Table 6.2: MC Move frequencies used in LaSSI simulations............................ 259
Table A.1: Scaffold mutations associated with disease whose phase behavior changes from that of WT. .......................................................... [320]
Table A.2: Known ligands and their effect on scaffold phase behavior. .............. [323]
Table A.3: Known ligands and their effect on scaffold phase behavior. .............. [326]
Acknowledgments

Firstly, I would like to thank my parents for loving me, supporting me, and always believing in me. Both modeled qualities that I continue to strive for. From my father, I received the spirit of duty – to fearlessly pursue your goals with every ounce of your being. From my mother, I received the spirit of perseverance – to always get back up smiling, no matter how painful the fall. They let their only child, of 17, go to a different continent so he could pursue his dream of studying physics. I know that it was not an easy decision to make for two people from Kashmiri Mohalla, and I am eternally grateful for the trust they continue to place in me. I would like to thank my paternal grandfather – he always called me Doctor Saahab... and so here we are. I would also like to thank my maternal grandmother for knitting me sweaters my whole existence, showing me that love is infinite and boundless, and that kindness and compassion are always the right response. As a very special shout out, I would like to thank Sir Yousaf, my high school mathematics teacher. His ardent interest, patience, and persistence provided me with a stable foundation to launch off of.

I would like to thank members of the WashU Physics Department. The culture of camaraderie has helped me stay mentally healthy, and I cherish my friendships with my colleagues. I thank Sarah Akin for being patiently vigilant and not letting me slip up with regards to all the formalities of being a graduate student. I also thank Henric for letting me rotate in his lab, and for sponsoring a conference trip when I had been in his lab for barely 2 months.
Furthermore, I would like to heartily thank my cohort: Kainen, Jesse, Natalia, John, Patrick, Yuewen. Be it staying up all night doing homework, or vigorously cheering each other at Karaoke, you have all made my time in St. Louis enjoyable, and unforgettable. I would especially like to thank Kainen for being one of my best friends, for being ever supportive, and for being so dependable.

I would like to thank the members, past and current, of the Pappu Lab. Each and every member I have interacted with has been warm, inviting, curious, and supportive. I thank Jeong-Mo for taking me under his wing when I joined the lab, and for laying down the foundations of what would end up becoming LaSSI. I thank Martin for sharing his deep knowledge of statistical mechanics and thermodynamics, and for being a great friend. I thank Alex for always answering my questions with rigorous detail no matter how simple the question, and of course to agreeing to be on this committee. I thank Mina for seriously and mindfully entertaining my ideas, no matter how absurd, and for the PS5. I thank Andrew for letting me come and bother him about biological questions, and about video games. I thank Matt for sharing his deep knowledge of nucleolar biology, and for always being patient in his explanations. I thank Kasia for making all the administrative issues in the lab a piece of cake, and for her honesty. I thank Xiangze for sharing his deep knowledge of computational techniques, and for playing tennis with me. I thank Sam for geeking out about applications of theoretical physics, and about sharing our love for food. I thank Ammon for sharing his deep knowledge about experimental techniques, and for inspiring me to be better at art. I thank Jared for always helping out, no matter what the ask, or what he is doing. Lastly, I am especially thankful to Kiersten. She has patiently mentored me the entire time I have been here, and has always taken time out to guide me. Her work ethic is nothing short of inspirational, and I am grateful to have her as a very close friend on top of all of that.
I would like to thank my PI and mentor, Rohit V. Pappu. His knowledge of the literature – both old and current – is awe inspiring. His expectations of unrelenting scientific rigor have made me a better scientist at every step in my journey in the lab. I especially appreciate his humility and candor in treating me like a scientific colleague who can challenge any claims he makes, and for taking my doubts and concerns seriously despite the vast gap in raw knowledge. Furthermore, that he and Natasha also treat me affectionately like their very own child has made the long years away from my homeland just that much more bearable – I am forever grateful to you both.

Being able to work with a vast network of collaborators has also been a privilege, and a boon. Without their active discussions, and profound experimental insights, my growth as a scientist would have certainly been stymied. I would like to thank the Rosen Lab, the Gladfelter Lab, the Alberti Lab, the Kriwacki Lab, the Knowles Lab, and the Hyman Lab. In particular, true to his name, Joy Kar has been wonderful to do science with, and I warmly look forward to doing more science with him. Lastly, I’d like to thank the rest of my committee. Anders for his patience and refined pedagogy, Shankar for his infectious excitement about biophysics, and James for his humility and honesty.

Finally, I would like to thank my partner and fiancée, Casey. For making me a better person, for being my home, for being my fellow goofball, for eating and enjoying all the food I cook, for passionately singing songs with me, and for being my best friend and closest confidant. Her support and love provide me with immense strength, and life’s colors are just that much more vibrant with her. I cannot wait to see where life takes us.

Furqan Dar

Washington University in Saint Louis
August 2022
Biomolecular condensates represent a new and ubiquitous class of membraneless organelles (MLOs) that are essential for healthy cellular functioning. The constituent molecules of such condensates span a vast bio-macromolecular gamut from intrinsically disordered regions and proteins (IDPs/IDRs), to RNA and RNA-binding proteins (RNPs), to polymerases and DNA etc. Apart from being part of the regular healthy cell cycle, these condensates are also implicated in many diseases, most notably progressive neurodegenerative diseases like Amyotrophic Lateral Sclerosis (ALS) and Huntington’s Disease (HD). Since the constituent molecules of these condensates span a broad range of length scales and modes of interaction, uncovering a unified framework that explains the underlying physical mechanisms for the formation of such condensates becomes important. While biological processes are decidedly out of equilibrium, equilibrium descriptions can still be useful, given the right constraints. Non-equilibrium descriptions tend to be mean-field-like, and we cannot access mechanistic details of the underlying interactions among molecules. Thus, making direct sense of the sequence encoded chemical specificity of different bio-macromolecules becomes a challenge. As such, an equilibrium mechanism for the formation of biomolecular condensates is that of
phase separation coupled to percolation (PSCP). Using the physics of associative polymers as a theoretical foundation, this thesis aims to uncover the molecular mechanisms and rules for biomolecular condensates formed via PSCP.

This thesis has two overarching goals. Firstly, to develop a computational approach that allows us to investigate PSCP at the “molecular” scale. Secondly, to use that computational approach to model and understand the workings of interesting and relevant biological systems. Along the way, with help from various experimental collaborators utilizing a plethora of techniques, we test our predictions, which we use to further refine our understanding. Towards the first goal, we have developed a flexible Monte-Carlo (MC) lattice polymer simulation engine: LAattice simulation engine for sticker and spacer interactions (LaSSI). Formulated as an extension to Bond-Fluctuation Models (BFMs), LaSSI allows for both the rigorous investigation of the equilibrium statistical properties of associative polymers, and also enables investigations of the phase transitions of such systems. We show that LaSSI can be used to generate full phase diagrams, including percolation, or gel, lines. Using an example linear multivalent protein (LMVP) system, we demonstrate the validity of LaSSI, including the assessment of finite-size effects in estimating a novel order parameter for detecting phase separation. We then extend to multiple and branched components using mimics of nucleolar proteins as the example system. Apart from the full phase diagram, and percolation boundary, we also show that the concept of a saturation concentration, $c_{\text{sat}}$, breaks down for condensates formed via heterotypic interactions.

Continuing, we use LaSSI to understand the modulation of biomolecular condensates. For a broad range of condensates, only a few components, called scaffolds, are necessary to drive the formation. Such scaffolds can be considered heteropolymers comprised of interacting motifs, called stickers, connected by solubility determining regions, or spacers. Using stickers-and-spacer representation in LaSSI, and the polyphasic linkage formalism, we have uncovered some
of the underlying physical principles of ligand-mediated modulation of the phase behavior of such scaffolds. For condensates that are primarily driven by heterotypically interacting scaffolds, ligands offer many different ways of modulation. On the one hand, monovalent ligands tend to only suppress phase separation whether they bind to stickers or spacers. On the other hand, divalent ligands produce a variety of effects. Ligands that interact with stickers tend to suppress the driving forces for condensation. Ligands that interact with spacers can enhance phase separation. While ligands that are bipartite, interacting with both stickers and spacers, can exhibit either suppression or enhancement, in a non-monotonic manner, depending on the asymmetry of the interaction strengths of the ligands between the stickers and spacers. Strikingly, even ligands that have minimal effects on $c_{\text{sat}}$ still modulate the internal structure of the condensates. Since LaSSI allows for accessing mechanistic details behind the modulation, our results provide a set of design principles for making bespoke ligands that can affect the phase behavior or internal organization of a condensate in a prescribed manner.

Moving on, we then pivot to understanding the structural regulation of a biomolecular condensate that is actively performing a function. Using some of the mitochondrial transcriptional machinery, our collaborators have been able to generate a minimal model of such a biomolecular condensate. Our collaborators show that these condensates are capable of transcribing RNA \textit{in-vitro}. Strikingly, the rates of RNA production are lower than well mixed systems, despite the increased local concentration of components. The components make heterogeneous multi-component condensates that seem to be kinetically and dynamically arrested. Using computational models that recapitulate the phase behavior of binary mixtures of the components, we then correctly predict that the condensates have spatially inhomogeneous organization of the underlying components. Furthermore, we show that the structure of the condensate is altered as more RNA is transcribed. We then predict
that this disruption of the condensate structure can be mitigated by the inclusion of RNPs, which are then confirmed by in-vivo assays performed by our collaborators. Importantly, we demonstrate that the vesicular structures seen in in-vitro experiments are out of equilibrium, and dynamically arrested phases. This is then verified experimentally where the order of addition of RNA produces very different organization of the condensates.

Lastly, we investigate the subtleties of systems that undergo PSCP. While Classical Nucleation Theory (CNT) predicts that the sub-saturated regime should be comprised mostly of monomers and exponentially small populations of oligomers, the physics of associative polymers predicts that this need not be the case. Indeed, our experimental collaborators show that the FET-family of proteins forms continuous heterogeneous distributions of oligomeric species, or clusters. Strikingly, small populations of mesoscale clusters, from 100 nm to 1000 nm, exist in a truly sub-saturated regime. Such clusters are quasi-spherical, thermodynamically stable, reversible and easily exchange material with solutions. We demonstrate not only that this is a general feature of the FET-family of proteins but that there is sequence specificity for the driving forces of cluster formation. We then further show that the interactions that lead to cluster formation and condensate formation can be decoupled using appropriate chemical solutes, or amino-acid substitutions. Furthermore, the data suggest that multiple length and energy scales are required to generate such mesoscale clusters. Indeed, using LaSSI simulations we show that homopolymers are not capable of generating mesoscale clusters but can only drive condensation in an all or none fashion. Heteropolymers, or associative polymers, are able to generate mesoscale clusters where the length asymmetry and interaction strength tunes the driving forces for cluster formation. The in-vivo implications are immense since most cellular scaffold concentrations are well within this subsaturated regime. That there are multiple components within the crowded environment of the cell makes an even stronger case for studying this phenomenon deeper where we could uncover the sequence
principles behind the coupling of percolation and phase separation, and what the biological function of such clusters might be. We thus find that equilibrium descriptions of biomolecular condensates still have much more to offer, and that the physics of associative polymers provides a stable foundation with which we can explore the molecular details of PSCP.
Chapter 1

Introduction

1.1 Preamble

This thesis focuses on developing flexible computational approaches that can be used to study the equilibrium statistical properties of associative polymers. We then use this tool to investigate the mechanistic details of how different bio-macromolecules can assemble into distinct macroscopic phases, and how such phases can be modulated. As such, we start with a general overview of biomolecular condensates. This overview is followed by taking a look at the types of molecules that can drive the formation of these condensates, which then serves as a foundation to find an appropriate conceptual framework. We then discuss the conceptual framework, that of associative polymers, to understand biomolecular condensates. Furthermore, we go over some possible ways to investigate the underlying physics of the formation of such condensates, and their limitations in describing the phenomenology in the larger context of biological processes. Lastly, the overall organization of the thesis is outlined.
1.2 Biomolecular Condensates

Biomolecular condensates have been identified as an exciting mechanism for enabling spatio-temporal organization in live cells. Where canonically we used to think that the membrane-bound organelles were the major way of segregating the biochemistry required to perform a particular function, there was always the discomfort of not being able to explain seemingly membraneless organelles (MLOs). One particularly important MLO, the nucleolus, was discovered in the 1830s. This thesis, then, aims to understand the equilibrium thermodynamic physical principles underlying the formation, modulation, and regulation of such MLOs, or biomolecular condensates.

While we, as a species, were aware of the nucleolus since the early 19th century, it was not until around 2009 that we started to understand the mechanism behind the formation of such MLOs. Clifford Brangwynne, working as a post-doctoral fellow for Anthony Hyman, paid careful attention to the *P Granules* in the germline cells of *C. Elegans*, [1]. By attaching fluorescent probes to the main constituents of these *P Granules*, PGL-1 or GLH-1, Brangwynne et. al., were able to carefully observe these granules under different physical conditions. Strikingly, these *P Granules* exhibited classical liquid-like behavior both in-vitro and in-vivo. With workable sizes around 1-4 µm, these *P Granules* could be extracted from cells at different stages in the cell-cycle. Bringing two droplets close to each other resulted in the droplets fusing together. By subjecting these droplets to shear-flows or shear-stresses, these droplets could be deformed and even be made to flow. These droplets also exhibited wetting behavior which changed depending on the substrate. They could use a sharp object to force the droplets to undergo fission. Of course, such tinkering was not just for fun. By carefully analyzing the curvature and the timescales of the droplets fusing, they could make estimates of the ratio between the surface tension, γ, and the viscosity η of these droplets. Furthermore, using
photobleaching experiments, they could then independently get a measure for the apparent viscosity, $\eta$, inside the droplets. With high viscosities, around $\eta \approx 1 \text{ Pa} \cdot \text{s}$, and low surface tensions, $\gamma \approx 1 \mu\text{N/m}$, these droplets were instantiations of highly deformable visco-elastic fluids. More importantly, given these liquid-like properties it was then hypothesized that the underlying mechanism was one we are all familiar with: condensation. In particular, liquid-liquid phase separation (LLPS) was suggested. In LLPS, much like a classical gas-liquid equilibrium, as the concentration of a particular molecule exceeds the so-called saturation concentration, $c_{\text{sat}}$, the solution becomes thermodynamically unstable. This instability is overcome by the separation of the system into two or more coexisting phases, or regions where the concentrations of particular molecules are higher or lower than the bulk concentration we start with. If we have a binary solution comprised of a solvent and solute, as the solute concentration is increased beyond $c_{\text{sat}}$, the solution separates into two phases: a dilute phase, and a dense phase (Figure 1.1).
Figure 1.1: Cartoon Depiction Of Phase Separation Of A Binary System (A) Solute concentration, $[c]_{\text{bulk}}$ vs. temperature, $T$, phase diagram for a binary system that undergoes phase separation. The shaded area represents the two-phase regime. Inside the two-phase regime, the system phase separates into coexisting dilute and dense phases, with respect to the solute. (B) One particular way to assess if two phases co-exist is to measure the solution concentration, $[c]_{\text{sol}}$, as a function of the bulk concentration, $[c]_{\text{bulk}}$. As we increase the total concentration $[c]_{\text{sol}} = [c]_{\text{bulk}}$ at first. As the bulk concentration rises above $c_{\text{sat}}$, we see that $[c]_{\text{sol}} = c_{\text{sat}}$ in the two-phase regime. Or, we see that $[c]_{\text{sol}}$ is saturated. Although most times impractical, in principle we get $[c]_{\text{sol}} = [c]_{\text{bulk}}$ as $[c]_{\text{bulk}}$ goes beyond $c_{\text{den}}$, where we find that the system is in the one phase regime yet again.

This exciting discovery by Brangwynne et. al., [1], inspired a series of insights for other biological systems and membraneless organelles. It turned out that the underlying mechanism for the formation of many MLOs was suggested to be LLPS. As opposed to a couple of decades ago, we now have a rather substantial, and growing, catalogue of MLOs which are presumably generated by LLPS. This catalogue has now grown, [2–11], to include Stress Granules (SGs), Cajal Bodies, U Bodies, certain membrane signalling clusters, etc. Furthermore, there is growing evidence showing that many progressive neurological diseases such as Amyotrophic Lateral Sclerosis (ALS), Huntington’s Disease (HD), Alzheimer’s Disease (AD) etc.
occur due to aberrant liquid-solid phase transitions of particular molecules: FUS, Htt-ex1, Amyloid-β, etc., [4, 12–31]. With a larger catalogue of MLOs and disease pathways, comes a deeper knowledge of the underlying mechanisms for the formation of biomolecular condensates. As we explore the physics behind the formation of such biomolecular condensates, we realize that rather than conventional LLPS, the more general concept of Phase Separation Coupled To Percolation (PSCP) is more appropriate, [32–39]. Indeed, proteins are not simple homopolymers made of the same building block, but are rather heteropolymers with many different interaction modes over many distinct length scales. How these interaction modes and length scales are coupled, and how these can be effectively tuned and modulated, represent the questions this thesis aims to shed light upon.

1.2.1 Constituents Of Biomolecular Condensates

As we try to make sense of the physical processes that drive the formation of biomolecular condensates, it is helpful to note the vast diversity of the molecules that comprise such bodies. Biomolecular condensates seem ubiquitous, [2, 4, 40]. They can be found within the cytoplasm, within the nucleoplasm, and in the extra-cellular matrix, [2, 4, 41]. We see that these condensates can be formed by proteins with large folded domains connected by linkers, [2, 6, 7]. Such systems represent heterotypically driven condensation, where the two functional groups reside on different molecules. While the folded domains end up as the primary drivers of the phase transition, the linkers determine the coupling between percolation and condensation. The amino-acid sequence of these linkers connecting the folded domains determine the type of coupled transition they undergo. On the one hand, if the linkers have high effective volume of solvation, the macromolecules only undergo a networking transition and form gel-like system spanning networks. On the other hand, if the effective volume of
Solvation is lower, the proteins also undergo condensation, forming highly connected micro-gel like droplets, [34, 42].

Similarly, these condensates can be formed by RNA, proteins, and RNA-Binding Proteins (RNPs), [2, 43, 44]. In one class of such condensates, the drivers of condensation are the canonical RNA Binding Domains (RBDs) and their cognate sites on the RNA(s). In particular, oligomerization due to Oligomerization Domains (ODs) results in higher total valency of the constituent RNPs which can dramatically increase the propensity to phase separate, [2, 39]. Interestingly, such ODs can also oppose condensation as shown by Seim et. al., [38].

Conversely, the drivers can be sticky motifs usually found within Low Complexity Domains (LCDs) and Intrinsically Disordered Regions (IDRs) of proteins, [2, 4, 14, 17, 45, 46]. Focusing on IDRs in particular, we now have a growing sense of the emergence of a so-called molecular grammar, [45], or sequence specific details, that determine whether an IDR can drive condensation. Here, we have homotypically driven condensation. In particular, interactions between aromatic residues like Tyrosine (Y), Phenylalanine (F) and Tryptophan (W) – the π-π interaction – can be the drivers of condensation for a set of LCDs, [45, 47–49]. Furthermore, cation-π interactions between the aromatic residues and the Arginine (R) can also be the primary drivers of condensation. The total valence, and the patterning of such residues, drastically affects the underlying driving forces, and the final networked structure of the condensate.

Lastly, and by far not exhaustively, for yet another class of biomolecular condensates, we can go back and think about the nucleolus. Here, DNA, RNA, and the interacting proteins can all come together to drive the formation of complex multi-phasic and multi-layered condensates, [50–59].
At the risk of seeming needlessly pedantic, proteins, RNA, and DNA are macro-molecules, [39, 42]. Furthermore, these macromolecules are often associative, in that there exist energetically favorable interactions between different molecules. If not the entire molecule, particular parts of the molecules can have high affinities for each other. Indeed, without such sticky interactions, proteins would not even adopt precise and stable folded conformations, let alone produce complex multi-subunit structures like the Ribosome. Conversely, when the stickiness is just right, reversible non-stoichiometric biomolecular condensates can form. How, and why, do multivalent macromolecules with sticky bits reversibly condense? Therefore, having a conceptual framework that can guide us in our investigations is paramount. This framework should preferably not confine us to a particular length scale, energy scale, or mode of interaction. In its generality we should be able to consistently and coherently think about biomolecular condensates driven by seemingly disparate mechanisms – little sticky bits on IDRs on the one hand, and giant folded domains tethered by linkers on the other. We explore precisely such a framework in the next section.

1.3 Phase Separation Coupled To Percolation: Theoretical Background

While many biological processes are far from equilibrium, with energy being consumed and new molecules being formed, equilibrium descriptions for the formation of biomolecular condensates are still powerful, [2, 33, 39]. As we make our way to PSCP, we shall consider phase separation and percolation independently. One particularly intuitive way to describe the demixing of polymer solutions was one pioneered by Flory and Huggins, [60–62], in the early 1940s. Thinking about Hildebrand’s regular solution theory, Flory made the intuitive leap to consider the physical consequences of solutions made up of multiple subunits, or monomers.
In other words, the solutions is comprised of *macromolecules*. Restricting ourselves to a lattice with a total of \( N \) lattice-sites, we can algebraically calculate the entropy of mixing. Consider two components, \( A \) and \( B \) with \( X \) and \( Y \) molecules respectively. Further, assume that \( A \) and \( B \) are made up of \( N_A \) and \( N_B \) monomers respectively. Assuming that the per-monomer volume, \( v \), is the same for the two components, we have \( N = X N_A + Y N_B \). The entropy of mixing of such a binary system containing two polymeric components is then:

\[
\Delta S_{\text{mix}} = -k_B \left[ \frac{\phi}{N_A} \log(\phi) + \frac{(1 - \phi)}{N_B} \log(1 - \phi) \right],
\]

(1.1)

where \( \phi = \frac{X N_A}{N} \) represents the *volume fraction* of component \( A \). The entropy of such systems *decreases* as the chain lengths of the components is increased, as shown in Figure 1.2.

![Figure 1.2](image_url)

**Figure 1.2:** The entropy of mixing, \( \Delta S_{\text{mix}} \), for a polymeric system as a function of polymer length. The total entropy of mixing decreases as the polymer gets longer. In effect, the longer a polymer, the easier it can be to keep it demixed if interactions are ignored.
Flory’s next intuitive leap was to consider systems where there are *interactions* between the species. In particular, he envisioned that in the mean-field, where we only care about *average* quantities, and where the connected nature of the polymer can be ignored, we can derive the Enthalpy of mixing, $\Delta H_{\text{mix}}$:

$$\frac{\Delta H_{\text{mix}}}{k_B T} = \chi \phi (1 - \phi), \quad (1.2)$$

where $\chi$, Flory’s parameter, represents the overall sum of polymer-polymer, polymer-solvent, and solvent-solvent interactions. We then finally have the free-energy density of mixing for such a system, by combining Equations (1.1) and (1.2). This gives us the Flory-Huggins (FH) free energy of mixing:

$$\frac{f}{k_B T} = \left[ \frac{\phi}{N_A} \log(\phi) + \frac{(1 - \phi)}{N_B} \log(1 - \phi) \right] + \chi \phi (1 - \phi), \quad (1.3)$$

As $\chi$ is decreased, which can result from favorable polymer-polymer interactions, or unfavorable polymer-solvent interactions, the free energy can become thermodynamically unstable, as shown in Figure 1.3.
As Flory was working on a theory that could describe the thermodynamics of polymer demixing, he was also developing a model that could predict the formation of polymer-gels, [63]. Here, the goal was to reliably predict the molecular size distributions of polymerization reactions where the functional groups were known \textit{a priori}. In particular, for a polymerization reaction, what would it take to generate a polymer component that is infinite, or rather of the order of the entire system. Almost concurrently, Stockmayer was pursuing the same questions, [64]. While Flory’s approach was based more on physical intuition, Stockmayer’s approach was based on explicit combinatoric calculations. Nonetheless, they produced consistent predictions. The amalgamation of the two we now know as the Flory-Stockmayer Theory of Gelation, [63, 64]. Given a macromolecule with $k$ functional groups, the percolation threshold, $p_c$, or the gelation point, $c_g$, is
As the number of functional groups is increased, the probability of forming a system spanning network goes up. Now, rather than considering the formation of chemical bonds, we consider transient and physical bonds between already polymeric systems.

We thus arrive at the theory of associative polymers, [42, 65–68]. Here, we combine both aspects of polymeric systems: the thermodynamics of FH-like demixing, and the thermodynamic consequences of transient physical bonds. Rather than a simple homopolymer, we now consider a heteropolymer. In particular, chemically specific associative motifs, or stickers, are connected by stretches of non-associative, or spacer, motifs. In this stickers-and-spacers framework, then, the spacers can be thought to determine the overall solution compatibility. The stickers, on the other hand, represent motifs that can form transient pairwise-only bonds. One particular formulation of the thermodynamics of such a system was introduced by Semenov & Rubinstein, [65, 67], which gives us the following free energy density:

\[
\frac{f}{k_B T} = \frac{\phi}{N} \ln \left( \frac{\phi}{\epsilon N} \right) + \frac{v}{2} \phi^2 + \frac{w}{6} \phi^3 + \frac{\phi}{l} \left[ \frac{p}{2} + \ln(1 - p) \right],
\]

where \( \phi \) represents volume fraction of the polymer as before; \( N \) the length of the polymer; \( v \) and \( w \) represent the 2-body and 3-body terms for the spacers, respectively; \( l \) the number of spacer monomers per chain; \( \epsilon \) represents the pairwise association energy for stickers; and \( p \) is the degree of conversion for the transient physical bonds formed by stickers. Thus, we have a thermodynamic formulation that explicitly accounts for transient connectivity of the system.
Given this free energy density, what new insights do we gain? The most striking prediction is that even for a system where the solution conditions are $\theta$-like, the free energy becomes unstable past the gel-point, $c_g$, as defined in Equation (1.4), provided the energy scale $\epsilon$ is strong enough. Therefore, for such systems we necessarily have that $c_{\text{sat}} < c_g$. Phase separation is thus coupled to percolation, and by tuning the spacers and stickers we can go from purely solution driven phase separated droplets, to systems that only percolate and form gels\(^1\). Chemically specific interaction motifs, be they folded domains, associative motifs on DNA or RNA, or individual amino-acids, can be considered stickers – the rest of the given macromolecule can be considered spacer-like. As we think about biological systems, it is good to remind ourselves that the identities of the stickers and spacers need not remain the same under different solution conditions. This theory, then, represents the foundational framework of the work presented here.

\section*{1.4 Numerically Investigating PSCP}

\subsection*{1.4.1 Field Theoretic Approaches}

As we have seen in the previous section, techniques based on mean-field and analytic approaches can be powerful descriptions of the possible phase behavior of associative polymers, and thus of biomolecular condensates, [42]. Even within the realm of equilibrium phenomenon, such analyses can provide legitimately testable hypotheses. Indeed, sticking to the stickers-and-spacers framework, the gel-point should scale inversely to the number density of stickers in a

\footnote{While the work presented in this thesis is related primarily to the equilibrium, and static, descriptions of polymeric systems, it should be noted that this theory also provides predictions for the dynamics of such solutions. Rather than being purely viscous fluids, we get visco-elastic fluids that behave elastically at shorter timescales, while they flow viscously over longer time scales.}
system. Wang et. al., [45], demonstrated that for the FET-family of proteins, \( c_{\text{sat}} \approx c_g \sim \frac{1}{n_R n_Y} \)

where \( n_R \) and \( n_Y \) are the number of Arginine and Tyrosine residues, respectively.

Yet, such formulations have an even more tantalizing advantage: probing non-equilibrium and active phenomena. There is growing consensus that biomolecular condensates cannot be described as simple newtonian fluids, [8, 36, 37, 69, 70]. Indeed, even the P Granules investigated by Brangwynne et. al., [1], were evidently *visco-elastic* fluids. As opposed to conventional fluids which can be effectively described by a simple viscosity, visco-elastic materials behave differently depending on the timescales at which they are probed. Having complex shear moduli, they behave elastically at shorter timescales, and viscously at longer timescales. An example of such a material that is commonly\(^2\) brought to bare is *Silly Putty™* which bounces when we throw it on the floor, yet effectively becomes a puddle if left out on a table. As suggested before, such visco-elasticity is predicted within the stickers-and-spacers framework, [62, 65]. The strength of sticker interactions can tune the timescales of making and breaking the transient bonds which, coupled to the solution conditions and polymer architecture, give rise to a range of visco-elastic behavior.

Furthermore, field-theoretic approaches also allow for the modeling of a different non-equilibrium phenomenon, namely that of particle production and degradation. Cellular concentrations of biomacromolecules vary greatly over time simply as part of the normal cell-cycle. Going back to the P Granules, [1], a concentration gradient formed by asymmetrical protein production in the growing embryo was essential ingredient for the robust formation of these P Granules in the right location. From mechanisms to regulate the sizes of the condensates formed, to active modulation by chemical reactions within the condensates themselves, field-theoretic approaches provide a plethora of deep insights, [71–73]. However, relating the sequence specificity of bio-macromolecules and the local chain connectivity of

---

\(^2\)Oobleck, or a lot of corn starch in water, preferably with an appealing dye, is another household example.
polymers, to the eventual predictions made becomes a non-trivial endeavor. The vast sequence
diversity found within bio-macromolecules, while having high chemical specificity, forces us
to think more deeply about explicit mechanistic details about how such molecules interact.
This, then, is the primary reason why the work presented in this thesis is based more on
explicit polymeric simulations.

1.4.2 Explicitly Modeling Macromolecules

The methods rooted in analytical, and mean-field, descriptions of PSCP cannot give us
mechanistic detail at the molecular level. Owing to the high degree of chemical specificity
of interactions, it becomes important to understand which of biomacromolecules are doing
what. One particularly successful and robust class for obtaining such detail can be broadly
categorized as Molecular Dynamics (MD) simulations, [74–83]. By explicitly modeling
every atom, solvent and solution ions included, we can brute-force numerically integrate the
equations of motion for the entire system. The interactions between all the atoms are usually
encoded in Force-Fields (FFs). These force-fields can range from simple Lennard-Jones (LJ)
interactions, to sophisticated multi-body interactions. The aim in adding more complicated
terms is usually to better reproduce high resolution experimental data - be it structural
or dynamical. While canonical MD force-fields have been fairly successful at tackling the
dynamics of folded proteins with a reasonably well defined energetic minimum, they do a poor
job in sampling Intrinsically Disordered Proteins (IDPs). Owing to their vast conformational
heterogeneity, [3, 25, 84], the energy landscape of such molecules lacks a well-defined energetic
minimum. Conventional MD techniques fail to enable large collective changes to parts of a
system. Here, by leveraging Monte-Carlo (MC) techniques, and implicit solvation models,
[85, 86], even the simulations of IDPs become tractable, at the single-chain level.
While all-atom simulations can provide very detailed mechanistic insights into how a particular bio-macromolecule interacts with itself and the solvent, investigating the phase behavior can become challenging rather quickly. Since phase transitions are emergent phenomenon, we need molecule numbers at least on the order of $10^3$-$10^4$. With each molecule potentially containing hundreds of residues, where each residue in turn contains dozens of individual atoms, the number of particles being simulated grows very quickly. Although advances in computer hardware and computer science have made such calculations possible, sweeping over the large parameter space required for inferring the underlying physical principles still remains intractable. Instead, we can reduce the overall complexity and Coarse-Grain (CG) the systems. By averaging over the interactions between predetermined groups of molecules together, we can generate a new type of effective particle. At length scales larger than the full collection of atoms, we can generate a pretty good approximation of how the collection of atoms will interact with another collection of atoms. We thus generate an effective mapping of the detailed interactions between the defined groups of atoms to a new interaction which reasonably approximates some predetermined measure of the interactions between these groups, [87]. Such CG simulations have been particularly successful in investigating the phase behavior of IDPs, [87–89]. However, losing the fine grained resolution of all-atom simulations means that the internal organization of condensates is harder to interpret, and that the models produced are usually not reliably transferable to other systems. Keeping this in mind, we realize that the power then lies in being able to compare different parameter sets for a given system, while keeping the models fairly general.

One further simplification can be made: we can discretize space and consider our CG polymers on a lattice. This discretization evidently produces its own challenges in the form of lattice artefacts since polymer conformations have also been discretized. This reduction in conformational space generally speeds up the calculations, and such lattice
representations are particularly amenable to more sophisticated MC techniques, [75, 83, 90]. Again, unlike MD simulations, MC simulations do not need to explicitly integrate the equations of motion describing the system. Instead, we sample a particular thermodynamic ensemble directly using sophisticated statistical techniques. This also means that we can no longer get detailed dynamical information about the processes being modeled, but instead only the equilibrium distributions for measured quantities. Yet, moving away from MD allows for dramatic configurational changes which need not even be physically realizable. This allows to circumvent some intrinsic kinetic barriers, enabling more efficient sampling of large collective changes of the polymeric systems.

Combined with simpler, and thus faster, modes of interactions, such CG lattice formulations allow for flexible development of phenomenological CG models, and larger sweeps over parameter space, [83, 89, 91–93]. Given some particular representation, and set of parameters that capture the desired phenomenology, we can easily investigate how this phenomenology changes as we change different parameters. Put simply, after finding a reasonable model, we can ask simple what if questions, e.g., what if this particular residue interacts with another residue differently, what if we add another component to the system, what if we add additional modes of interactions. What are the changes to the phase behavior, to the connectivity, and to the internal organization of condensates formed. Indeed, the work presented in this thesis has this flavor, and we shall explore this in detail in Chapters 2 and 3. As such, by using tools that allow for the explicit simulations of large numbers of molecules we can track the mechanistic details behind the emergent phenomena – from the sequence encoded statistical mechanics, to the eventual thermodynamics.
1.5 Scope Of The Thesis

The main scope of the work presented in this thesis is two-fold. Firstly, we focus on the development of an open-source computational tool designed for investigating the physics of associative polymers explicitly on a lattice. Secondly, we then deploy this tool to investigate a number of questions with a focus on generating mechanistic insights behind the phenomenon in question. By leveraging explicit coarse-grained representations of the relevant biological macromolecules that capture a particular phenomenology, we get to ask how the emergent properties of the system change as we perturb the underlying parameters of the models.

Chapters 2 and 3 are primarily focused on the development and physical basis for the open-source computational LaSSI. LaSSI, or LAttice Simulation Engine For Sticker And Spacer Interactions, represents a general and flexible Monte-Carlo (MC) tool that can be used to investigate the physics of associative polymers on a lattice. These polymers can be coarse-grained (CG) representations of biological molecules, e.g. IDPs, RNA, and ligands. Since we use the MC method, and because these polymers exist on a lattice, LaSSI allows for easy scalability where we can include $10^3$-$10^4$ molecules. This then allows us to investigate the phase behavior of such polymer systems where we can explicitly track the density and networking transitions, while also being able to probe the mechanistic details and internal structures. Focusing on the internal details of LaSSI, Chapter 2 describes the move-set employed, the interactions, and the micro-reversible biased sampling. Chapter 3 then demonstrates using LaSSI, and that we generate rigorous and physically relevant results.

As we saw earlier, biomolecular condensates in-vivo are rarely comprised of just a few molecular species. Instead, these condensates are comprised of many different molecular species. Where sometimes it is possible to identify so-called scaffolds which drive the formation of condensates, the rest of the molecular species can act as modulators and regulators. In
Chapter 4 we focus on modulating the phase behavior of such scaffolds using ligands. Ligands represent molecular species that cannot drive the formation of the condensates, and that they interact with specific binding sites on the scaffolds. Connecting these results with the general framework of polyphasic linkage, [94], we show that there is great diversity in how ligands can modulate the phase behavior of scaffolds. We find that ligands monomeric ligands in general tend to increase $c_{\text{sat}}$, while for dimeric ligands even if $c_{\text{sat}}$ remains unchanged, the internal structure of condensates is affected. Furthermore, using relationships generated purely from theoretical considerations, we show that using the partition coefficient (PC) of ligands alone convolves multiple contributions and cannot be trivially used to determine whether a ligand stabilizes or destabilizes the condensates. Hopefully, such insights can guide us, as a community, to better design and identify bespoke modulators of phase behavior.

The functional consequences of condensates is still an open question in the field, [2, 11, 14, 95]. In Chapter 5, we take a deeper look at the transcriptional machinery in Mitochondria. The mt-Nucleoid, believed to form due to phase separation, [53, 54], serves as a viable model for condensates that perform a function: transcription. Using minimal in-vitro reconstitutions of this transcriptional machinery, we systematically go over the phase behavior of the components. We see that indeed the in-vitro condensates transcribe mt-RNA, but the rates of RNA production are much lower than a well mixed system. A LaSSI model that qualitatively matches phase behavior of binary mixtures is used to make predictions about the equilibrium structure of the condensates formed by the full mixture. These predictions lead us to believe that the condensate structures observed in-vitro during active transcription are non-equilibrium dynamically arrested phases. Indeed, by performing order of addition experiments, and measuring the intrinsic mobilities of the components using FRAP support this hypothesis. and that by having suitable RNA processing molecules, we can circumvent this structural change.
Since biomolecular condensates represent the existence of a density phase transition, the canonical expectation generated by Classical Nucleation Theory posits that the subsaturated solutions for such systems are comprised primarily of monomers, and exponentially small numbers of oligomers. Yet, as pointed out above, biomolecular condensates most likely form due to phase separation coupled to percolation. Therefore, in Chapter 6 we take a deeper look at the oligomeric composition of subsaturated solutions. In particular, we focus on the FET-family of proteins where. We find that subsaturated solutions of such proteins have continuous and heterogeneous distributions of oligomers or clusters. Strikingly, we observe thermodynamically stable meso-scale clusters which are distinct from the condensates that eventually form above saturation. These clusters exhibit sequence-specificity and solute-sensitivity. Using insights generated from experimental data, we develop a minimal model that can generate heterogeneous cluster distributions. This then leads us to hypothesize that multiple length scales or energy scales are required to observe subsaturated clustering.

Lastly, Chapter 7 is a summary of the work presented in this thesis. We reiterate the main results produced, highlight the major insights pertaining the phenomena in question, and suggest possible avenues for further investigations – some of which we have already started thinking about in earnest.
1.6 References


Chapter 2

LaSSI: Lattice Simulation Engine For Sticker And Spacer Interactions

2.1 Preamble

This chapter is based primarily on the following paper: Choi JM*, Dar F*, Pappu RV (2019), LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology, 15*(10): e1007028. (*Co-first authors*) [1]. J.M.C., F.D., and R.V.P designed the research; J.M.C. and F.D. performed the research; F.D performed the data analyses and visualizations; J.M.C, F.D., and R.V.P. wrote the paper.

Finally, since LaSSI (https://github.com/Pappulab/LASSI) is an actively developed software, major changes have been made since the publication of the original manuscript. In particular, new Monte-Carlo moves have been implemented and as such we explicitly note new moves as they appear in the chapter.
2.2 Introduction

Biomolecular condensates organize cellular matter into non-stoichiometric assemblies of proteins and nucleic acids [2]. Prominent condensates include nuclear bodies [3] such as nucleoli, nuclear speckles [4, 5], and germline granules [2, 6, 7]. Condensates also form in the cytoplasm. These include stress granules [8], membrane-anchored signaling clusters [9, 10], and bodies in post-synaptic zones [11]. All of these condensates seem to share key features: (i) they range in size from a few hundred nanometers to tens of microns [2, 3, 12]; (ii) they are multi-component entities comprising hundreds of distinct types of proteins and nucleic acids; (iii) and of the hundreds of different types of molecules that make up condensates, a small number are essential for the formation of condensates [2, 13]. The simplest feature that distinguishes proteins that are drivers of biomolecular condensates is the valence of interaction domains, or motifs, that can participate in non-covalent crosslinks [2, 13–15].

Biomolecular condensates can form and dissolve in an all-or-none manner [3, 12, 16]. The reversible formation and dissolution of condensates can be controlled by the concentrations of multivalent proteins that drive the formation of condensates; in simple two-components systems comprising of macromolecules and solvent, condensates form when macromolecular concentrations cross macromolecule-specific threshold values known as saturation concentrations [16]. The transitions that characterize condensate formation bear the hallmarks of a sharp density transition, leading to the formation of a dense phase that is in equilibrium with a dilute phase. This type of transition, known more broadly as phase separation, generates two or more coexisting phases. Working under equilibrium conditions, the chemical potential of each macromolecule is equalized across the phase boundaries [16]. Phase separation is reversible where we can achieve this reversibility by: (i) making changes to concentrations of the driver macromolecules [10, 17]; (ii) making changes to solution conditions, like temperature,
that alter the effective interaction strengths among driver molecules [18–21]; (iii) altering saturation concentrations through ligand binding - a phenomenon known as polyphasic linkage [22–25] (iv) via biological regulation such as post-translational modifications of proteins [9, 13, 26].

Recent studies have focused on uncovering the defining features of proteins [14, 16, 18–20, 27–43] and RNA molecules [44–46] that drive phase transitions. Protein and RNA molecules that drive phase transitions are biological instantiations of associative polymers [47] characterized by a stickers-and-spacers architecture [48]. Stickers generally contribute to a hierarchy of specific pairwise and higher-order interactions that are either isotropic or anisotropic, while spacers control the concentration-dependent inhomogeneities in the densities of stickers around one another. Stickers can be hot spots or sectors [49] on the surfaces of folded proteins [16, 32] or short linear motifs within intrinsically disordered regions (IDRs) [16, 27, 50]. Spacers are typically IDRs that contribute through their sequence-specific effective solvation volumes to the interplay between density transitions (phase separation) and networking transitions that are better known as percolation [31, 32]. Spacers can also be folded domains that are akin to uniformly reactive colloids. Proteins can be mapped onto this stickers-and-spacers architecture as linear multivalent proteins, branched multivalent proteins, or some combination of the two [14, 16].

Simple biological two-component systems comprise the solvent (which includes all components of the aqueous milieu) and a multivalent protein / RNA molecule. For fixed solution conditions, one can generate phase diagrams [28] as a function of protein concentration, the valence of stickers, the affinities of stickers, the sequence-specific effective solvation volumes of spacers, and the lengths / stiffness of spacers. The phase diagram can be investigated by keeping the valence of stickers, the lengths of spacers, and effective solvation volumes of spacers fixed while varying the concentration of stickers and the affinities between stickers [32]. Increasing protein
concentrations will enable larger density fluctuations, and above the saturation concentration, designated as \( c_{\text{sat}} \), the density inhomogeneities lead to separation of the system into coexisting phases. The concentration of multivalent proteins in the dilute and dense phases will be denoted as \( c_{\text{sat}} \) and \( c_{\text{den}} \), respectively. For a given bulk concentration \( c_{\text{bulk}} \) that lies between \( c_{\text{sat}} \) and \( c_{\text{den}} \), the fraction of molecules within each of the coexisting phases is governed by the lever rule [51]:

\[
c_{\text{bulk}} = f_{\text{sat}} c_{\text{sat}} + f_{\text{den}} c_{\text{den}},
\]

where \( f_{\text{sat}} \) and \( f_{\text{den}} \) represent the volume fractions of the dilute and dense phases respectively. As a reminder, the lever rule is simply stating mass-balance explicitly: if we add up all the molecules from each phase, we should get back the total amount of molecules in the system to begin with.

Stickers also form reversible physical crosslinks and these crosslinks generate networks of inter-connected proteins. The number of proteins within the largest network of the system grows continuously as the protein concentration increases. Above a concentration threshold known as the percolation threshold and designated as \( c_{\text{perc}} \), the single largest network spans the entire system and this phenomenon is called percolation [52–54]. If the percolated networks have the rheological properties of viscoelastic fluids, the fluids are referred to as network fluids [16, 55].

Theory [18, 27, 28, 30, 37, 56–62] and computations [31, 32, 46, 63–71] have important roles to play in modeling and describing the phase behavior of multivalent protein and RNA molecules. Theories provide analytical routes to explain experimental observations and to make testable predictions. On the other hand, simulations work around many of the
simplifying assumptions that are needed to make theories analytically tractable. In doing so, they provide numerical routes to enable comparative assessments across different systems; they help in making testable predictions about phenomenology through what if calculations targeted toward specific systems; and they pave the way for designing systems with bespoke phase behavior.

Phase transitions are collective phenomena that involve highly cooperative transitions of large numbers of multivalent polymers. The collective interactions that drive phase transitions are captured in terms of a small number of order parameters that are similar across disparate systems and represent a generic coarse-graining of the underlying system that defines parameters such as the correlation length and the sizes of cooperative units. Accordingly, practical considerations of computational tractability and rigorous considerations of identifying the relevant collective coordinates mandate the use of coarse-grained models for simulations of phase transitions driven by multivalent protein and RNA molecules. We focus here on multivalent proteins, although the methods we describe are readily adaptable to RNA molecules as well.

Coarse-graining, an essential aspect of making simulations of large numbers of multivalent proteins a tractable proposition, comes in different flavors [72]. For simplicity, we divide considerations that go into the development of a suitable coarse-grained model into three categories (Fig 2). These are (1) the type of model, (2) the types of interactions among the entities in the simulation, and (3) parameterization of the interaction potentials for the model of interest.

Two distinct choices for the type of model are the choice between simulations being performed using lattice models versus off-lattice models. In either space, one or all of the molecules can be represented explicitly using architectures that represent coarse-grained mappings of
the protein of interest. Next, the interactions among the units that make up each protein can be modeled as being isotropic or anisotropic. This is true of simulations where proteins of interest are modeled explicitly. In contrast, numerical instantiations of field theoretic models model can also be brought to bear where only a single chain is modeled explicitly [63, 73]. The remaining protein and solvent molecules are modeled as fields whose fluctuations are concentration dependent [74]. The effects of all other molecules influence the phase behavior of the explicitly modeled single chain through interactions of the chain with the field. Finally, the choice of interaction potentials is the bedrock of every simulation. The functional forms and parameters for potentials can be derived using phenomenological considerations intended to enable calculations of the “what if” variety-an approach that is common practice in statistical and polymer physics. One can also obtain system-specific parameters using information gleaned from atomistic simulations of smaller-scale facsimiles of the system of interest. These system-specific parameters are derivable using force matching methods pioneered by Voth and coworkers [75–80] or by prescribing a functional form for the potential that describes interactions in the coarse-grained space and employs machine learning methods to derive the relevant parameters [77, 81]. Finally, one can adopt approaches similar to the parameterization of molecular mechanics force-fields and develop a single transferable model that should be applicable to a large number of disparate systems.
Figure 2.1: As discussed in the text, the choice of a coarse-grained model has at least three ingredients. These include the type of conformational space (lattice or off-lattice), the nature of the interactions among entities that are represented in the coarse-grained description (isotropic, anisotropic or fluctuating fields), and the parameterization approach. LaSSI, as described here, is based on a lattice model that uses anisotropic interactions and a phenomenological model. Lastly, as LaSSI was further developed, we also included isotropic interactions which are discussed in later chapters as they come up.

Different coarse-grained simulations represent different combinations of model, interaction type, and parameterization. Two illustrative examples for deriving coarse-grained models for simulations of phase behavior of multivalent proteins come from the works of Ruff et al. [81] and Dignon et al. [67, 68]. Ruff et al. show how one can generate off-lattice models, of bespoke resolutions and learned parameters for isotropic potentials derived using machine learning that leverages information gleaned from atomistic simulations of individual proteins and protein oligomers. Dignon et al. also use an off-lattice model based on isotropic potentials whose parameters are designed to be transferable across disparate intrinsically disordered proteins.
It is worth emphasizing that at this juncture, there is no valid reason to stipulate that one combination of approaches for deriving a coarse-grained model is superior to another combination. As noted by Das et al. [70, 71], all models have distinct strengths and limitations. However, for specific applications, some methods afford quantifiable computational advantages over others. In our case, we are interested in uncovering conceptual nuances of phase diagrams for multicomponent systems that comprise multivalent proteins characterized by anisotropic interactions among domains / motifs. Furthermore, we want to have the flexibility of being able to generate minimal and simple models for such a broad range of molecular architectures, and interaction modes.

As noted above, these systems can be mapped onto a stickers-and-spacers architecture. The questions we are interested in answering pertain to the order parameters that describe phase behavior, the impact of chain connectivity and spacer effective solvation volumes on phase behavior, and the determinants of the shapes of phase diagrams of multicomponent systems where phase transitions are driven by heterotypic and homotypic interactions. In this context, it is noteworthy that lattice models have been adapted to model phase transitions for systems comprising different numbers of multivalent protein and RNA molecules [31–33, 43, 82–84].

In this chapter, we provide a formal description of the design of a general and flexible simulation platform that can be used to generate system-specific lattice models for simulating phase transitions of multivalent proteins. This simulation engine, known as LaSSI for LAttice simulation engine for Sticker and Spacer Interactions, formalizes the approaches that have been developed and deployed in recent studies [31–33, 82, 83]. LaSSI combines a lattice model with anisotropic, pairwise-exclusive, interactions among stickers. We describe the design of LaSSI, focusing first on the overall structure of the model, the Monte Carlo (MC) sampling, and their justification for generic multivalent proteins.
In the next chapter, using two specific examples of linear and branched multivalent protein systems, we illustrate the deployment of LaSSI to two biologically relevant systems.

2.3 LaSSI Model

The structure of LASSI is inspired by the bond fluctuation model (BFM) for lattice polymers [85]. This is a general lattice model for simulations designed to extract equilibrium conformational distributions and dynamical attributes of polymers in dilute solutions as well as dense melts. There are two versions of the BFM viz., the Carmesin-Kremer BFM or CK-BFM [86] and the Shaffer BFM, or S-BFM [85]. Both models are based on the use of simple cubic lattices, which discretizes the conformational space for polymers.

In the CK-BFM [86], each repeating unit or monomer within a polymer is modeled as a 3-dimensional cube where the 8 corners of the cube occupy lattice sites and bond vectors connect pairs of monomers. Overlap of monomers is associated with an energetic penalty, and each bond vector can have up to 108 distinct directions. The choice of bond vector set encodes the geometry of the polymer and places constraints on the bond lengths and bond angles. All other interactions are governed by the inter-monomer potentials, and evolution of the system through conformational space is driven by changes to the overall potential energy.

In contrast, the S-BFM places each monomer on a single lattice site. Covalently bonded monomers are connected by bonds that are constrained to be of three types, leading to chains that have bonds of length 1, $\sqrt{2}$ or $\sqrt{3}$ in units of lattice size. Monte Carlo moves with suitable acceptance criteria can be designed for both types of BFMs. The simulations are used to generate equilibrium conformational distributions of lattice polymers in either dilute or dense phases. The move sets control the overall polymer dynamics and the acceptance of different types of moves and the calculation of correlation functions allows one to compute
dynamical quantities for lattice polymers [85]. If we were to use either of the established BFM\s without modification, then each amino acid residue would be modeled as a monomer, and such an approach would be useful when the identities of stickers and spacers remain ambiguous. This approach underlies a different simulation engine known as PIMMS [46].

2.3.1 Extension To Bond Fluctuation Models

The current LaSSI model is a generalization of the bond fluctuation model prescribed by the S-BFM where we also adopt some features from CK-BFM, [85–87]. Given a CG description of a particular set of macromolecules, polymers of arbitrary shape are described by a set of non-overlapping monomers that occupy 3-dimensional cubic lattice sites. Depending on the coarse-graining, the monomers on the lattice might not represent the monomers of the particular macromolecule. Monomers are linked via **phantom** bonds such that the bonds act as tethers and not as rigid rods. This also means that while monomers are not allowed to overlap, bonds can overlap. Any bond length, $l$, can be defined provided that $l$ is smaller than $L_{\text{min}}/2$, where $L_{\text{min}}$ is the smallest lattice dimension. A bond length of $l$ means that adjacent monomers must be within $l\sqrt{3}$ of each other as shown in Figure 2.2.
Figure 2.2: 2D representation of a 2-component system within the LaSSI framework. Since the bonds act as tethers bonds freely cross over each other, and that the distance between neighbouring monomers fluctuates. Lastly, the dashed lines represent a physical bond due to the anisotropic, or rotational, interaction between the two monomers.

Inter-monomer (sticker-sticker, sticker-spacer, and spacer-spacer) interactions are modeled as contact-based pairwise interactions. A sticker monomer can bind to another sticker monomer that occupies an adjacent lattice site with an interaction energy that depends on the types of both monomers. Monomers are considered to be adjacent to one another if they are within a lattice distance of $\sqrt{3}$. By this criterion, each lattice site occupied by a sticker monomer will have 26 adjacent lattice sites. This is reminiscent of the interaction geometry of a CK-BFM for each monomer. In the current implementation of LaSSI, the anisotropic interactions...
are mutually exclusive, implying that a sticker cannot bind simultaneously with more than one other sticker, even though there are 26 adjacent sites where other stickers could be present. This combination of the geometry of the interaction sites per monomer and the single occupancy constraint leads to effective anisotropic interactions between stickers. This feature is unique to LaSSI and it is not incorporated in other variants of BFMs; this allows us to deploy LaSSI for modeling heteropolymeric systems.

In the context of LaSSI, we note that stickers are distinguished by their ability to participate in anisotropic or isotropic interactions. In contrast, explicitly modeled spacer sites only participate in isotropic interactions with other spacer or sticker sites. Furthermore, the interaction strengths involving spacers are typically weaker than those involving stickers. However, it is worth emphasizing that these distinctions only matter inasmuch as LaSSI allows us to capture a numerical instantiation of the stickers-and-spacers model. For simplicity, one might simply think of LaSSI as a model that has sites that are differentiated by whether or not they can involve themselves in anisotropic (or isotropic) interactions, and by the comparative magnitudes of site-site interaction strengths.

2.4 Move Set

Our goal is to compute architecture-specific phase diagrams for systems comprising of one or more types of linear or branched multivalent proteins. This requires a simulation strategy that enables the sampling of the full spectrum of coexisting densities and networked states for multivalent proteins. Both the polymer conformations, and the anisotropic sticker-sticker interactions need to be considered. Accordingly, the conformations of randomly initialized systems of proteins on a simple cubic lattice are sampled via a series of Markov Chain Monte Carlo (MCMC) moves that are designed to ensure efficient sampling of changes in protein
density and networking while maintaining microscopic reversibility. Thus, we need a set of moves that can efficiently sample both the dilute phase, which can correspond to the dilute or semi-dilute regime, and the dense phase, which can correspond to highly networked polymer-melt-like conditions. Below is a collection of moves that we have developed and deployed that aim to simulate such systems.

2.4.1 Biased Monte Carlo Sampling

Given a system in configuration $i$, in the Canonical, or $TVN$, ensemble, the Boltzmann weight corresponding to this state is

$$P(i) = \frac{\exp[-\beta E(i)]}{\mathcal{Z}}, \quad \text{(2.2)}$$

where $E_i$ is the total energy of the system, and $\mathcal{Z}$ is the canonical partition function. A MCMC move that transitions the system from configuration $i$ to configuration $f$ represents a flow in configuration space. Under equilibrium conditions, this flow must satisfy microscopic detailed balance:

$$P(i)\pi(i \to f) = \pi(f \to i)P(f), \quad \text{(2.3)}$$

where $\pi(i \to f)$ represents the transition probability from $i$ to $f$, and $\pi(f \to i)$ the reverse. Furthermore, we note that the transition probability can be further decomposed as such:

$$\pi(i \to f) = \text{acc}(i \to f)\alpha(i \to f), \quad \text{(2.4)}$$
where $\text{acc}(i \to f)$ is the probability of accepted moves made from $i$ to $f$, and $\alpha(i \to f)$ represents the underlying matrix of the Markov chain, or the move proposal distribution. Plugging (2.4) into (2.3), we get

$$P(i)\text{acc}(i \to f)\alpha(i \to f) = P(f)\text{acc}(f \to i)\alpha(f \to i). \quad (2.5)$$

Rearranging the terms to get the ratio of the move acceptances, we have:

$$\frac{\text{acc}(i \to f)}{\text{acc}(f \to i)} = \frac{P(f)\alpha(f \to i)}{P(i)\alpha(i \to f)}. \quad (2.6)$$

One particular solution to this ratio of acceptances is of course the one by Metropolis-Hastings (MH), [88]:

$$\text{acc}(i \to f) = \min \left\{ 1, \frac{P(f)\alpha(f \to i)}{P(i)\alpha(i \to f)} \right\}, \quad (2.7)$$

where the more frequent form assumes that $\alpha(i \to f) = \alpha(f \to i)$, giving

$$\text{acc}(i \to f) = \min \{1, \exp [-\beta(E(f) - E(i))]\}. \quad (2.8)$$

In LaSSI, we want the underlying MCMC moves to also efficiently sample the sticker-sticker interactions. We, thus, use Rosenbluth Sampling [89–91] to enhance physical bond interactions, while still satisfying detailed balance. In particular, we note that for a given system, the total energy can be split into position-dependent interactions, and sticker-sticker interactions:

$$E(i) = E_{i,\text{pos}} + E_{i,\text{sti}}. \quad (2.9)$$
We modify the move proposal distributions as such, \[89, 90\]:

\[
\alpha(i \rightarrow f) = \frac{E_{i,sti}}{W_i},
\]

where

\[
W_i = \sum_k \exp \left[ -\beta E_{a,sti} \right],
\]

is the Rosenbluth factor \[90\] which represents the Boltzmann distribution over \( k \) randomly sampled sticker orientations at the location \( i \). Therefore, plugging everything back into MH acceptance criterion from Equation (2.7), we have:

\[
\text{acc}(i \rightarrow f) = \min \left\{ 1, \frac{W_f}{W_i} \exp \left[ -\beta (E_{f,pos} - E_{i,pos}) \right] \right\}.
\]

This modified acceptance criterion is used in multiple moves, and Sub-section \ref{subsec:2.4.3} goes into more detail about how we actually calculate the Rosenbluth weights.

\section*{2.4.2 Sticker Rotation Moves}

This move serves as the basic unit of sampling the anisotropic interactions between monomers. Monomers can be considered \textit{associated} when they are engaged in an anisotropic interaction, or \textit{dissociated} otherwise. Since the anisotropic interaction, and thus the physical bonds, are stochastically modeled, this move is both conceptually and computationally simple.
We randomly pick a monomer from the system and attempt to change its binding state. If the monomer cannot interact anisotropically, the move is immediately rejected. Else, we search neighboring lattice sites for possible binding partners. In the case of LaSSI, the accessible volume for anisotropic interactions is within the cube of 1 lattice-site around a bead – see Figure 2.3 for a 2D representation. Therefore, we have $k = 3 \times 3 \times 3 - 1 = 26$ possible neighboring sites. Since $k$ is not large, we choose to exhaustively go through all 26 possible neighboring sites.

Figure 2.3: In 2D there are $k = 3 \times 3 - 1 = 8$ neighboring sites (yellow box). Eligible candidates are monomers that have a non-zero interaction energy with the selected monomer (middle orange bead). In this figure, orange monomers only interact with blue monomers, and thus we have 3 possible candidates. The final binding state is selected using the usual Metropolis-Hastings criterion, where the non-bound state is also included.
Suppose we end up with \( n \leq k \) possible candidates. Assume further that the interaction energies are all the same value, \( \epsilon \). Then, proposing a new state is equivalent to uniformly sampling \([0, n]\). We draw a random integer \( r \sim U(0, n) \), where \( n \) is included. If \( r = 0 \), we propose that the bond be broken. In the more general case where not all interaction energies are the same, if \( r > 0 \), we then sample the Boltzmann Distribution generated from the list of possible candidates. Finally, when a candidate is picked, we use the usual Metropolis-Hastings (MH) criterion for acceptance or rejection.

### 2.4.3 Local Moves

This move serves as the basic unit of local movement for beads. A randomly chosen monomer is randomly displaced from its starting position, and is only allowed to move to empty lattice sites which still satisfy the linker constraints for that monomer.

Suppose we have picked a monomer, indexed \( n \). This monomer is moved from \( \vec{r}_i \) to \( \vec{r}_f = \vec{r}_i + \Delta \vec{r} \), if there is no site overlap with another monomer and linker constraints are satisfied. We construct \( \Delta \vec{r} \) by uniformly sampling \( U(-l, l) \) for each component, where \( l \) is the linker-length of the between monomers \( n \) and \( n - 1 \) (or between \( n \) and \( n + 1 \) is \( n = 0 \)). This constrains \( |\Delta \vec{r}| \leq l\sqrt{3} \), as shown in Figure 2.4. This can result in moves corresponding to crankshaft and kink jump moves if the selected monomer is in the interior of a molecule, and end rotation moves if end monomers are selected, [92, 93].
Figure 2.4: 2D representation of a local move. For a given randomly selected monomer, a new location is proposed by sampling $\pm l (= 2)$ lattice sites in each coordinate (brown box). In this example, the proposed lattice site is empty (black dot) and thus the monomer is translated to the new site. The numbers of interacting neighbors are calculated at the old and proposed sites. These are then used to construct a modified Metropolis-Hastings acceptance criterion which determines if the move is accepted or rejected.

Focusing on the anisotropic interaction, we start by explicitly stating the energy difference between the two states. Suppose that this monomer $n$ goes from being bonded to $l$ to being bonded to $m$. Further, let $n$ have type $q$, $l$ have type $x$, and $m$ have type $y$. Then,

$$E_{i,\text{rot}} - E_{f,\text{rot}} = \epsilon_{qx} - \epsilon_{qy}. \quad (2.13)$$

Plugging this into Equation (2.6), we now get:
\[
\frac{\text{acc}(i \rightarrow f)}{\text{acc}(f \rightarrow i)} = \frac{\exp(-\beta \epsilon_{qx}) W_{i;n}}{\exp(-\beta \epsilon_{qy}) W_{f;n}},
\]

where we still keep \( n \) around as a reminder that the Rosenbluth weights are for that particular monomer. Here, we now can choose to further decompose the Rosenbluth weights as a sum:

\[
W_{i;n} = W_{i;n}^{(a)} + W_{i;n}^{(d)},
\]

where of (\( a \)) is the associated part and (\( b \)) is the dissociated part. Suppose that \( n \) has a set of neighbors, \( \Theta \). The sum representing the associated part can be written out explicitly as:

\[
W_{i;n}^{(a)} = \sum_{t \in \Theta} \exp(-\beta \epsilon_{qk_t}),
\]

where \( k_t \) represents the type for monomer \( t \). Now, to make the calculations simpler, to help us build intuition, we suppose further that there is only one energy scale in the system, \( \epsilon \). Then, the explicit sum in Equation (2.16) is simply the number of neighbors for monomer \( n \) scaled by the Boltzmann weight for that energy – namely:

\[
W_{i;n}^{(a)} = N_i n e^{-\beta \epsilon},
\]

where \( N_{i;n} \) is the number of neighbors in state \( i \). For the dissociated part we set \( W_{i;n}^{(d)} = 1 \), which strongly biases the system towards associated states. Plugging all this into the modified Metropolis-Hastings acceptance term from Equation (2.8), we have:

\[
\text{acc}(i \rightarrow f) = \min \left\{ 1, \frac{N_{f;n} + 1}{N_{i;n} + 1} \exp \left[ -\beta (E_{i,\text{pos}} - E_{f,\text{pos}}) \right] \right\} .
\]
Finally, we see that acceptance ratio is now scaled in such a way where states with more neighbors have higher chances of being accepted. This scheme ensures detailed balance while correctly biasing the system to sample bonded configurations, increasing sampling efficiency. This procedure for the bias for a monomer also serves as the basic block for the more complicated moves that follow.

2.4.4 Co-Local Moves

This move was not in the original LaSSI manuscript, [1]. This move attempts to move a pair of beads that are engaged in a physical bond - see Figure 2.5 for a 2D representation. We first pick a random bead in the system. If the bead is not currently associated, the move is rejected. Assuming that we now have a pair of monomers, \(a\) and \(b\), that are engaged in a physical bond, or are associated, we attempt to displace the two monomers. A random displacement vector, \(\Delta \vec{r}\), with radius 2 is generated in a way similar to the local move (see Sub-section 2.4.3). If the proposed vector results in steric clash for either monomer, or if the linker constraints for either monomer are unsatisfied, the move is rejected. Otherwise, the monomers are displaced and the usual MH acceptance criterion is used if there are additional interactions in the system. Otherwise, the move is readily accepted. Since this move is only concerned with displacement, we do not use any biasing scheme.
2.4.5 Slithering Snake Moves

The slithering snake move is a Monte-Carlo realization of reptation as was first conceived by deGennes, [94]. As local monomer density increases, it gets harder and harder for local-like moves to generate large configurational changes. Furthermore, at high densities, such as ones that dense phases might adopt, linear polymers tend to travel along tubes [95, 96]. Therefore, this move aims to mimic those dynamics, albeit statically.

We start by picking one end of a randomly selected chain and propose to move the first monomer to a new position. Like the local move in Sub-section 2.4.3, we propose a trial location for the selected end of the chain. Steric clash for the end-monomer results in move rejection. Otherwise, the end-monomer is moved to the new location. Suppose that the end-monomer has index $i$. Then, monomer $i+1$ takes the place of monomer $i$, $i+2$ the place of $i+1$, and so on – see Figure 2.6 for a 2D representation. This move relies on an inherent
symmetry of chain molecules which are homopolymer-like where linker lengths between each monomer are the same and if one were to swap the linkers between any pair of monomers, the chain remains topologically invariant. Therefore, this move is rejected if a proposed molecule has heterogeneous bond lengths, or is a branched molecule.

Figure 2.6: 2D representation of a slithering snake move. For a given randomly selected chain where all linker lengths are the same, one end is randomly chose. In this example, the top end of the chain is selected. A random location is selected within a given radius (brown box). In this example, the chosen location neither results in steric clash, nor does it break the linker length constraints of the chain. As such, the top monomer is placed at the new location, and every monomer in the chain takes the place of the monomer ahead of it. Note that the move is anisotropically biased, which results in the formation of new physical bonds. A modified MH acceptance criterion is used to determine if the move is accepted or rejected.

This move is anisotropically biased for every monomer in the chain. Since the anisotropic states of monomers are mutually independent, the Rosenbluth factors can be computed for
each monomer independently as in Equation (2.18). Then, we calculate the total Rosenbluth weight by taking the product of all individual weights. Using the same notation as Subsection 2.4.3, we have

\[ W(i) = \prod_n W(i; n) = \prod_n (N_{i;n} + 1), \tag{2.19} \]

where the product is taken over the appropriate monomers in the chosen chain. As in Subsection 2.4.3, for ease of notation, we have assumed that all interaction energies are \(\epsilon\), and thus we can simplify the Rosenbluth factors a little. The MH acceptance criterion becomes

\[ \text{acc}(i \rightarrow f) = \min \left\{ 1, \prod_n (N_{f;n} + 1) e^{-\beta (E_{f,\text{pos}} - E_{i,\text{pos}})} \right\}. \tag{2.20} \]

The inclusion of the bias for every interacting monomer, rather than just the end monomers, is to emulate how a real transiently bonded polymer would slither along its contours when densities are high.

### 2.4.6 Multi Local Moves

*This move was not in the original LaSSI manuscript, [1]*. This move is designed to generate larger conformational changes than the local move within the interior of molecules by moving multiple monomers. We take advantage of the fact that bonds, be they physical or covalent, are allowed to overlap in LaSSI.

A particular monomer is randomly chosen, along with all of its covalently bonded monomers. We then remove all monomers from the lattice and attempt a local move for each of the monomers, see Figure 2.7 for a 2D representation. Each of the monomers is placed back
on the lattice within a 2-lattice site radius of the old position. Steric clash for any of the monomers results in immediate rejection of the move. If there is no steric clash for any of the monomers, we then check if the molecular topology constraints are satisfied. If any of the linker constraints are broken, the move is rejected.

Figure 2.7: 2D representation of a multi-local move. In this example, the orange monomer in the middle is picked. We start by removing it, and all monomers covalently bonded to it (orange monomers) from the lattice. Then, we place each monomer back in a random location around the old position. In this example, the new proposed locations do not result in steric clash and the linker constraints for each monomer are also satisfied. Furthermore, since this move is anisotropically biased, we see that the monomers can form new bonds or break older ones. This move can result in significant configurational changes in the interiors of polymers, even if the local density is fairly high, since bonds are allowed to overlap.

This move is also anisotropically biased, as discussed above Sub-section 2.4.5. The Rosenbluth weights are calculated for every monomer that is randomly displaced, as described in Sub-section 2.4.5. The modified MH acceptance criterion is then used to determine whether to accept the move or reject it.
2.4.7 Translation Moves

Up till now the moves have primarily focused on equilibrating dense configurations of the system and by construction do not result in large translations of the center-of-mass (COM) of the molecules. Therefore, we need moves that can translate the COM of molecules on the order of the simulation box, which will be explained in the current and next few subsections. The chain translation move is designed to move single chains while potentially also forming new bonds at the proposed location if possible candidates are nearby. If there are many intramolecular physical bonds for a particular chain, this move can also provide an opportunity for new bonds to form.

This move attempts to translate a chain from $\vec{r}_i$ to $\vec{r}_f = \vec{r}_i + \Delta \vec{r}$, see Figure 2.8 for a 2D representation. We construct $\Delta \vec{r}$ by uniformly sampling $[-L/4, L/4]$ for each component, where $L$ is the box-size for that component. For cubic boxes, $|\Delta \vec{r}| \leq \frac{\sqrt{3}}{4} L$, where $L$ is the box size. After $\Delta \vec{r}$ is constructed, we attempt to translate the chain. Steric clash for any of the monomers within the chain results in immediate move rejection. Like the slithering snake move described above we apply an anisotropical bias. Each monomer that has anistropic interactions in the molecule is biased, and the Rosenbluth factors are calculated as in Equation (2.19), and the acceptance criterion is modified as shown in Equation (2.20). This modified MH acceptance criterion is then finally used. This move not only results in large displacements for single chains but also correctly biases the system to efficiently sample bond configurations.
Figure 2.8: 2D representation of a chain translation move. For a given randomly selected chain, we attempt to displace the chain on the order of the simulation box size. In this example, orange monomers interact with blue monomers. The proposed displacement does not result in steric clash, and the chain is translated to the new location. The anisotropical bias is applied to every monomer in the chain, and therefore new physical bonds can form at the new location. Finally, we use the modified MH acceptance criterion to decide whether to accept or reject the move.

2.4.8 Old Cluster Translation Moves

Cluster translation moves translate multiple chains that belong to the same cluster, which is defined as the set of unique chains connected via anisotropic interactions. A proposed move only results in a translation, and the move is readily accepted if there is no steric clash. This move might seem unnecessary as this move simply moves clusters around, but we realize that once a physical bond has formed between two molecules, it is highly unlikely for the other types of moves to translate the COM’s of clusters closer together. This is demonstrated in a later section.
Cluster moves need to be defined very carefully so as to not break micro-reversibility. If a cluster is defined as nearest neighbors, then a proposed cluster move might change the number of clusters in the system – necessarily breaking micro-reversibility (see Sub-section 2.4.9). Due to the way physical bonds are modeled in LaSSI, the cluster definition is intuitive, and the cluster move is micro-reversible by construction. Since no new physical bonds are created at the proposed location, the cluster remains invariant under translation and the move is accepted.

2.4.9 New Cluster Translation Moves

As LaSSI evolved to include both heterogeneity of interaction energies, and the inclusion of contact-like overlap energies, we saw the need to implement new types of cluster translation moves. Rather than simply translating the cluster that a random chain belongs to, we now have two clustering criteria, and two variants for each criterion, resulting in four different moves:

1. Proximity Small Cluster
2. Proximity Second Largest Cluster
3. Anisotropic Small Cluster
4. Anisotropic Second Largest Cluster

For the Small variants, we attempt to move clusters that are smaller than half the number of chains in the system, \( N_T/2 \). We start by selecting a random chain, indexed as \( m \). We then calculate the cluster that \( m \) belongs to. If \( m \) is a monomer, the move is rejected. As a reminder, a cluster is defined as the unique set of connected chains where for Proximity-based clusters, a bond is defined as being within \( \sqrt{3} \) of a monomer, and for Anisotropic-based
clustering, monomers engaged in anisotropic interactions are considered bonded. If the size of the cluster is larger than $N_T/2$, the move is rejected outright.

For the Second Largest variants, we start by calculating the full cluster decomposition of the system corresponding to either criterion. We then select the second largest cluster from this list. If there are multiple second largest clusters, then we randomly select one of the second largest clusters. If the system is composed of monomers, this move is rejected.

As before, the goal of these moves is to translate clusters by $\Delta \vec{r}$. We construct $\Delta \vec{r}$ by uniformly sampling $[-L/4, L/4]$ for each component where $L$ is the box-size for that component. For cubic boxes, $|\Delta \vec{r}| \leq \frac{\sqrt{3}}{4}L$, where $L$ is the box size. We then attempt to translate the given cluster, where steric clash of any of the chains in the cluster results in immediate rejection of the move. Otherwise, the clustering definition determines the next step.

For the Anisotropic-based variants, since cluster translations do not alter the clusters, we readily accept the move. For the Proximity-based clusters, we have to take extra care to ensure micro-reversibility. In particular, we recalculate the clustering and make sure that the clusters remain the same after translation. If not, we reject the move. While this might seem counter-intuitive, we remind ourselves that the reverse move should always be possible. If cluster translation results in a new cluster, as shown in Figure 2.9, then if we start from the proposed configuration, our cluster would be larger, and thus
Care must be taken when defining cluster translation moves. A *proximity* based cluster is translated to a proposed location. The cluster is now neighbored by a new monomer, colored orange. Now, for the backwards move, we get a different cluster than the one we started with. This makes it impossible for the backwards move to return the system back to the old state.

### 2.4.10 Branched Rotation Moves

This move is designed to sample the configurations of branched molecules in the system. By performing simple 3D-rotation operations on branched molecules, we can generate substantially different configurations for those molecules. Since we are on a lattice, only the subset of rotations that produce integer values can be used. Assuming that our lattice is a simple cube with side $L$, we have rotations of $\pi/4$, $\pi/2$, $\pi$, $3\pi/2$, and $2\pi$ along any of the three lattice dimensions.

Given a randomly selected branched molecule, we perform a 3D-rotation operation on the whole molecule, using the center as the origin for the transformation. We randomly select a rotation operation to perform. We use the central bead of the molecule as an anchor-point, and perform the rotation operation on all other monomers of the polymer, see Figure 2.10 for a 2D demonstration.
Figure 2.10: 2D representation of a branched rotation move. Given a randomly selected branched molecule, we attempt to rotate the entire molecule where the central monomer (green bead) is used as the origin. In this example, we rotate the molecule by $\pi/2$, or 90°, around the axis out of the page. The proposed rotation does not result in steric clash for any monomers, and we perform the rotation. Since this move is anisotropically biased, the rotated monomers can form new physical bonds. A modified MH acceptance criterion is then used to determine whether to reject or accept the move.

Steric clash results in the rejection of the move immediately. To also efficiently sample physical bonds, an anisotropical bias similar to the slithering snake move’s is employed. Here, Rosenbluth weights are calculated, as described in Sub-section 2.4.5, for every bead except the anchor. With these weights, a modified MH acceptance is used to determine whether to accept or reject the move.

### 2.4.11 Pivot Moves

*This move was not in the original LaSSI manuscript, [1].* Continuing with thinking about applying simple 3D-rotation operations to collections of monomers, we have the Pivot Move. This move is especially effective at sampling dilute phase configurations for long polymers[91,
where application of rotation operations on parts of the chain can dramatically alter
the conformation.

In particular, for a given randomly chosen polymer in the system, we pick a monomer in
the chain, indexed \( m \), which acts as an anchor. Given this anchor we divide the chain into
two parts. We then attempt a 3D-rotation operation, much like in Sub-section 2.4.10, to
the shorter subsection of the chain with the anchor as the origin, see Figure 2.11 for a
2D representation). Steric clash results in immediate rejection of the move. Furthermore,
this move is also anisotropically biased to improve the sampling of bonds, as discussed in
Sub-section 2.4.5. Here, the bias is applied for all monomers that are moved due to the
3D-rotation. Then, a modified MH acceptance criterion is used to determine whether to
accept or reject the move.

![Figure 2.11: 2D representation of a pivot move. Given a randomly selected polymer, we
select an anchor monomer (green bead). We then select a rotation operation to perform on
the shorter section of the chain, given the anchor as the origin. In this example, we pick a
\( \pi/2 \), or 90\(^\circ\), rotation along the axis of the page. Since the rotation does not result in steric
clash, the monomers are moved to their new locations. Since the move is anisotropically
biased for those monomers, they can also form new bonds at those locations. The modified
MH acceptance criterion is finally used to determine whether to accept or reject the move.](image)
2.4.12 Double Pivot Moves

Lastly, we have the double-pivot move. As stated before, one key advantage of MC simulations over MD simulations is that the changes to the system between time-steps need not be physical. The double-pivot move is another move that leverages this flexibility. The double pivot move swaps a part of a chain with the corresponding part of another chain of the same type (Figure 2.12). The purpose of this move is to engender large configurational changes in dense polymer melt-like systems, which the dense phases are likely to be. If the density is high enough and the polymers become entangled, local moves behave like slithering-snake moves and polymers are restricted to travel along tubes [95, 96] around each other. For two chains close enough to each other that the bonds between two monomers can be swapped, then such a move should result in a large configurational change for both chains.
Figure 2.12: For a randomly selected monomer, a $2 \times 2 \times 2$ cube around the monomer is searched for appropriate bridging candidates (brown box), where an appropriate bridging candidate is the next monomer from a different chain, is within a linker length of the selected monomer as shown by the dashed line connecting $i_m$ and $(i + 1)_n$. Furthermore, the distance between $(i + 1)_m$ and $i_n$ must also be within a linker length as depicted by the upper dashed line. A list of all possible candidates is calculated and then a randomly chosen candidate is used to break and remake covalent bonds. This results in a large conformational change for both polymers. If the selected polymer is not linear, the move is rejected outright.

We start by randomly selecting a monomer, denoted as $i_m$, where $m$ is the index of the chain, and $i$ is the index of the monomer within that chain. The move is rejected if the molecule is not linear. A search is performed, within a set distance, around $i_m$ for monomers from the same type of chain. Specifically, we look for $(i + 1)_n$, where $n \neq m$ is the index of another chain. Therefore, we are searching for monomers on different chains that are the next intra-chain index. We set the radius of the search to be $R_{DP} = \min\{3, l\}$, where $l$ is the
linker length for \(i_m\) and \((i + 1)_m\). Any monomer within the search radius trivially satisfies the linker constraints of \(i_m\), and we store all such monomers and chains as potential candidates.

From this set of candidate chains, we further check if the distance between \((i + 1)_m\) and \(i_n\) satisfy their respective linker constraints as well. All candidates that do not satisfy these constraints are removed, and we end up with \(N_F\) candidates for the forward move. To ensure microscopic reversibility\(^3\), we also calculate the number of backwards candidates, \(N_B\), using the same procedure, but on \(i_n\) as the base monomer. A random candidate is then selected from the \(N_F\) forward ones where each candidate is uniformly probable, and the chains are partially swapped. Since LaSSI does not have any bonded interaction energies, the total energy of the system does not change due to the change in chain bond configurations. Lastly, a modified Metropolis-Hastings acceptance criterion is used:

\[
\text{acc}(i \to f) = \min \left\{ 1, \frac{N_F}{N_B} \right\},
\]

\(^3\)One possible way to avoid the double calculations involved in ensuring micro-reversibility is to have the move fail as soon as we have multiple candidates.
2.5 References


46. Boeynaems, S., Holehouse, A. S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P. S., Pappu, R. V. & Gitler, A. D. Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex...


Chapter 3

Using LaSSI To Investigate The Phase Behavior Of Multivalent Proteins

3.1 Preamble

This chapter is based primarily on the following paper: Choi JM*, Dar F*, Pappu RV (2019), LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology*, 15(10): e1007028. (*Co-first authors) [1]. J.M.C., F.D., and R.V.P designed the research; J.M.C. and F.D. performed the research; F.D performed the data analyses and visualizations; J.M.C, F.D., and R.V.P. wrote the paper.

3.2 Introduction

In this chapter we demonstrate the use of LaSSI by applying it to study two archetypal systems that are known to undergo phase separation [2–4]. The systems are instantiations of
linear and branched multivalent protein systems, respectively. The simulation results obtained
for linear multivalent proteins illustrate how phase diagrams are generated when protein
concentration (at a fixed stoichiometry) and temperature are the independent variables. In
the second example that includes a branched multivalent protein and a linear peptide, the
temperature is fixed, and the concentrations of the individual components are varied. The
simulation parameters for both systems are summarized in Table 3.1 below. For each system,
we conducted 5 independent simulations, each of which consists of $5 \times 10^8$ MC steps after
$5 \times 10^6$ equilibration steps. The data were collected over the last half of the trajectories at a
frequency of $5 \times 10^5$ steps.

<table>
<thead>
<tr>
<th>Bead Notations</th>
<th>FUS-Like System (Figure 3.1)</th>
<th>N130 + rpL5 System (Figure 3.9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Of Stickers $s_i$</td>
<td>$s_A = 5, s_B = 5$</td>
<td>$s_A = 10, s_B = 5$</td>
</tr>
<tr>
<td>Linker Length $l_{ij}$ (in lattice units)</td>
<td>$l_{AN} = 1, l_{BN} = 1$</td>
<td>$l_{AN} = 1, l_{BN} = 3$</td>
</tr>
<tr>
<td></td>
<td>$l_{NN} = 4$</td>
<td>$l_{NN} = 3$</td>
</tr>
<tr>
<td>Position-Dependent Energy $E_{pos}(r_1, r_2)$</td>
<td>$\infty$, if $r_1 = r_2$</td>
<td>$\infty$, if $r_1 = r_2$</td>
</tr>
<tr>
<td></td>
<td>0, otherwise</td>
<td>0, otherwise</td>
</tr>
<tr>
<td>Pairwise Interaction Energy $\epsilon_{ij}$ (in temperature units)</td>
<td>$\epsilon_{AB} = -3, \epsilon_{ii} = 0, \epsilon_{iN} = 0$</td>
<td>$\epsilon_{AB} = -3, \epsilon_{ii} = 0, \epsilon_{iN} = 0$</td>
</tr>
</tbody>
</table>

Table 3.1: Simulation parameters for system descriptions.

As we describe the two archetypal systems, we also go over the calculations of relevant
order parameters to detect phase separation using pair distributions, how finite-size artefacts
can affect these distributions, how we estimate the percolation-lines, why cluster moves are important for systems that undergo phase separation, and how we can explicitly construct a 2-component phase diagrams with the derived order parameters.

3.2.1 Building Intuition For Setting Move Frequencies

The frequencies of the different move sets for simulations of the two systems considered are summarized in Table 3.2 below. Considerations that go into the design of move sets include the achievement of converged equilibrium distributions, with maximal computational efficiency, for each bulk concentration. The structure of each move set serves as a guide for selecting an optimal set of frequencies. This leads to a set of heuristics that are as follows: (i) in the cluster move we pick a random chain from the system, perform a networking analysis on that chain, and then propose a displacement of the cluster. As the cluster size grows it is more likely that a randomly picked chain will be part of the largest cluster which itself will result in a steric clash after the proposed move. Therefore, the frequency of the cluster move should be low, if not the lowest, in the entire set. (ii) In the translational move, we pick a random chain from the system for translation; as the size of the largest cluster increases it becomes less likely for a proposed translation move to be accepted. However, unlike the cluster move the translation move is rotationally biased and thus results in new interactions being formed. Hence, translational moves enable single-molecule to cluster-surface interactions. Therefore, this move should be proposed more often than the cluster move, although not as often as rotational or local moves. (iii) The rotation move is computationally inexpensive and it enables the switching of physical bonds and should thus be proposed fairly frequently. (iv) Similarly, local moves and slithering snake moves are also rotationally biased, and they help with the local rearrangements of physical bonds. Local moves are the primary route to enable local conformational changes, and to enable local physical bond rearrangements. Therefore,
local moves should be proposed most frequently. The slithering snake move is particularly
effective because it allows for large local physical bond rearrangements in dense configurations.
Thus, this move should also be proposed frequently, less so than local moves but more so
than translation moves. Note that in a system where some molecules are non-linear or have
heterogeneous linker lengths, the frequency would need to be higher since the move is rejected
if an incorrect molecule is picked at random. (v) The double pivot move allows for large-scale
changes to conformations within dense configurations and accordingly, this move should
be proposed more frequently than both cluster and translation moves. One can track the
acceptance ratios of each move over a very rough initial sweep across the relevant system
parameters. Moves that are always rejected do not enable any changes in configuration and
only add computational costs. Therefore, the frequency for that particular move should be
lowered. This is especially the case for the cluster move in high-density systems.

<table>
<thead>
<tr>
<th>Move</th>
<th>Linear System</th>
<th>Branched System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Translation</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chain Translation</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Sticker Rotation</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Local</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Slithering Snake</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Double-Pivot</td>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3.2: Move frequencies according to different types for linear and branched systems. The frequencies are normalized with respect to the lowest frequency move.
3.3 Linear Multivalent Proteins

3.3.1 FUS As An Archetypal Linear Multivalent Protein

Wang et al. [2] recently uncovered the molecular grammar that contributes to the driving forces for phase separation of the protein FUS and a family of related proteins. They showed that, to first order, $c_{\text{sat}} \propto (s_Y s_R)^{-1}$, where $c_{\text{sat}}$ is the measured saturation concentration of the FUS family proteins and $s_Y$ and $s_R$ are the number of tyrosine (Tyr) and arginine (Arg) residues, respectively. In FUS and other proteins of similar architectures, the Tyr residues are located primarily within the N-terminal disordered prion-like domain (PLD), whereas the Arg residues are located primarily within the partially disordered C-terminal RNA binding domain (RBD).

Using mutagenesis experiments, Wang et al. established that Tyr and Arg residues are the stickers in the FUS family proteins. Accordingly, the zeroth-order stickers and spacers representation used to model FUS in LaSSI comprises of two parts: An N-terminal mimic of the PLD encompassing Tyr residues as stickers and a C-terminal mimic of the RBD that encompasses Arg residues as stickers. Wang et al. also measured $c_{\text{sat}}$ for a 1:1 mixture of independent PLDs and RBDs interacting in \textit{trans}. The $c_{\text{sat}}$ for this system is approximately twice that of the $c_{\text{sat}}$ for full-length FUS. Given the block co-polymeric architecture of FUS, we denote the PLD and RBD as $A_n$ and $B_n$, respectively for A and B-blocks of valence n. The model system of PLDs and RBDs interacting in \textit{trans} is denoted as $A_n + B_n$ (Figure 3.1a), whereas the system mimicking full-length FUS where the stickers can interact in \textit{cis} and in \textit{trans} is denoted as $A_n - B_n$ (Figure 3.1b).
Figure 3.1: Coarse-grained representation of an archetypal linear multivalent system. (a, b) Cartoons to depict the $A_n + B_n$ and $A_n - B_n$ systems, respectively. Different colors of beads denote different species of stickers. Note that $A_n - B_n$ can be simply considered as $A_n + B_n$ where the two different sections of the proteins are joined together. In this chapter, we consider FUS akin to a di-block co-polymer where the species of stickers are segregated to the two halves of the protein, while the linkers and spacers are the same throughout. (c) Linker lengths involved in the architecture (see also Table 3.1). Each sticker has a neighboring spacer bead that is 1 lattice site apart whereas the neighboring spacer beads are 4 lattice sites apart. This means that consecutive stickers are 6 lattice sites apart and also that the linkers connecting the two have a positive effective solvation volume.

Within $A_n$ and $B_n$ blocks, spacers provide a uniform spacing of six lattice sites between stickers along the chain. We model spacers using a hybrid approach whereby a neutral spacer monomer is attached to each sticker with spacing of a single lattice site (Figure 3.1c). This choice was made to provide a distinction between $A_n - B_n$ and $A_n + B_n$. Accordingly, the relative concentration of neutral beads will be higher in $A_n - B_n$ when compared to $A_n + B_n$. This allows us to study linker-mediated differences between the driving forces for phase separation for $A_n - B_n$ vs. $A_n + B_n$. The move set frequencies are summarized in Table 3.2.
### 3.3.2 Detecting Phase Transitions Using Distribution Functions

Along with developing a suitable and flexible software platform in the form of LaSSI, we also wanted to develop order parameters that can help us *detect* phase transitions. In order to detect the onset of phase separation, we can calculate excess chemical potentials using the Widom particle insertion method \[5\] and equalize these chemical potentials across distinct phases. This process requires a priori knowledge of the densities of both phases. An efficient variant of this approach, based on fast Fourier transforms, was recently developed and deployed by Qin and Zhou \[6\]. They demonstrated their method for calculations of liquid-liquid coexistence curves for a patchy colloid model of $\gamma$II-cyrstallin.

Given that LaSSI simulations are based on a lattice, and that we want a more general way to detect phase separation, we use pair-distribution functions. The pair-distribution is constructed by integrating over all degrees of freedom except pairs from a partition function as:

\[
P^{(2)}(r) = \int \exp[-\beta E\{\mathbf{r}^N\}] \, d\mathbf{r}_3 \cdots d\mathbf{r}_N,
\]

and serves as a rigorous thermodynamic and structural measure of density inhomogeneities. Intuitively, we are asking how *pairs* of objects are arranged around each other, given that we average over all the other objects in the system, and that further we also average over all possible pairs. In particular, if a system undergoes phase separation into two distinct phases, we necessarily get density inhomogeneities. With this pair distribution function, we can calculate the *radial distribution function* (rdf) by normalization using an appropriate prior distribution.
However, normalization of the pair distribution function requires some care. The system contains polymer molecules and using a prior distribution that assumes an ideal gas of the chain monomers to normalize the pair distribution function is problematic because it does not accurately capture the effects of non-idealities due to chain connectivity. We leverage the efficient sampling of polymer fluids in LaSSI, and obtain suitable prior distributions by simulating the system of interest in the absence of sticker-sticker interactions – generating an ideal pair distribution, $P_0^{(2)}(r)$, given the polymer architectures.

The pair distribution function $P^{(2)}(r)$ quantifies the equilibrium distribution of distances between pairs of chain monomers, where $r$ is the inter-monomer distance (see Figure 3.2). Likewise, $P_0^{(2)}(r)$ quantifies the same distribution of distances but without any interactions, and serves as the normalizing prior distribution (see Figure 3.2). Thus, we construct a new rdf:

$$
\tilde{g}(r) \equiv \frac{P_0^{(2)}(r)}{P^{(2)}(r)}. 
$$ (3.2)

The function $\tilde{g}(r)$ is a direct measure of the local density of the protein of interest. Since LaSSI uses periodic boundary conditions, the maximal inter-monomer distance is $\frac{\sqrt{3}}{2} L$. Given this normalized $\tilde{g}(r)$, we note that if the system has short-range ordering as in a canonical liquid, the radial distribution function will oscillate around unity but eventually approach one as $r \to \infty$. Conversely, if the system undergoes a density transition, $\tilde{g}(r)$ will have two distinct spatial regimes (Figure 3.2 (b)): for small $r$, $\tilde{g}(r)$ will be characterized by a tall and broad peak such that $\tilde{g}(r) > 1$ until $\tilde{g}(r)$ intersects the $\tilde{g}(r) = 1$ line; this region corresponds to the dense phase and we shall denote the value of $r$ at this intersection to be $r \equiv r_b$. For $r > r_b$, $\tilde{g}(r)$ will be between 0 and 1, and for lattices that are large enough to avoid finite size artefacts, $\tilde{g}(r)$ will converge to a value lower than one and this corresponds to the density in
the dilute phase region. Furthermore, \( \tilde{g}(r) \) can be used to estimate the densities within the dense and dilute phases.

Figure 3.2: The data shown are obtained from 5 independent simulations for the \( A_n - B_n \) system with total protein concentration \( c \approx 7 \times 10^{-5} \text{ voxels}^{-1} \) and reduced temperature \( T^* = 0.383 \). Error bars indicate standard deviations. (a) Pair distribution functions \( P^{(2)}(r) \) and \( P_0^{(2)}(r) \), where the former is from the interacting system and the latter from the non-interacting system with chain connectivity (prior pair distribution function). Note that \( P^{(2)}(r) \) shows two peaks, the first of which indicates dense phase formation. (b) Radial distribution function \( \tilde{g}(r) \). This captures the droplet formation by a sharp and broad peak in the beginning. The inset shows \( r_b \) where \( \tilde{g}(r) \) intersects the line corresponding to \( \tilde{g}(r) = 1 \), delineating between the dense and solution phases. The global density inhomogeneity measure, \( \overline{\rho} \), is obtained by integration of absolute deviation of \( \tilde{g}(r) \) from 1.

To quantify a measure of global density in homogeneity we introduce a simple measure, \( \overline{\rho} \), which is calculated as follows:

\[
\overline{\rho} \equiv \left( \frac{1}{L} \right) \int_0^{\sqrt{3}L} |\tilde{g}(r) - 1|V(r/L) \, dr, \quad (3.3)
\]
where \( V(x) \) is the volume element for a cubic lattice with side \( L \), given periodic boundary conditions, for a normalized radial distance \( x = r/L \). Following [7], we can use an analytical expression for this volume element:

\[
V(x) = \begin{cases} 
4\pi x^2, & \text{if } 0 < x \leq 1/2 \\ 
2\pi x(3 - 4x), & \text{if } 1/2 < x \leq \sqrt{2}/2 \\ 
2x(3\pi - 12f_1(x)f_2(x)), & \text{if } \sqrt{2}/2 < x \leq \sqrt{3}/2 
\end{cases}
\]  

(3.4)

and where

\[
f_1(x) = \arctan \left( \sqrt{4x^2 - 1} \right) 
\]

(3.5)

\[
f_2(x) = 8x \arctan \left( \frac{2x(4x^2 - 3)}{(4x^2 - 3)\sqrt{4x^2 - 1}} \right) 
\]

(3.6)

If \( \bar{\rho} \approx 0 \), the global density inhomogeneity in the system is small and this will be characteristic of a single homogeneous phase dominating the simulation volume. As \( \bar{\rho} \) increases beyond 0, the system accommodates density inhomogeneities. We construct coexistence curves using a cutoff value of \( \bar{\rho} = 0.025 \), which is universal to all systems, to delineate between a homogeneous system, and one that has undergone phase separation.

### 3.3.3 Quantitative Assessment Of Finite Size Effects

The pair distribution function is central to our calculation of density inhomogeneities and constructing coexistence curves for a system of multivalent proteins simulated using LaSSI. At the start of this section we emphasized the importance of including \( 10^3 - 10^4 \) distinct
molecules within the simulation cell in order to avoid finite size artefacts. Prior to presenting
detailed results that mimic specific systems, we present an analysis of finite size effects that
we will confront if the requisite numbers of molecules are not included in the simulations. The
data we present are for simulations of mimics of the protein FUS, specifically the \( A_n + B_n \)

system introduced in the results section. The phenomenological mapping of this protein
architecture onto a cubic lattice is discussed at the start of the results section. Here, we
present an analysis that makes a crucial technical point about finite size effects.

First, we start with simulations for ideal polymers. The data shown in Figure 3.3 plots
the pair distribution function \( P^{(2)}_0(r) \) extracted for simulations of ideal models of FUS-like
proteins. Results are shown for simulations that use 20 chain molecules of \( A_n + B_n \) as an
example of a small system. These results are compared to those from simulations with 100,
200, 1000, 1500, 2000, 3000, and 4000 \( A_n + B_n \) molecules, respectively. The pair distribution
functions have a self-similar character and this is revealed by plotting \( P^{(2)}_0(r) \) for all of the
simulations, where \( r^* \) is the reduced distance that accounts for the fact that for a similar
concentration, the simulation cells are made larger (higher values of \( L \), which is the box
size) as the numbers of molecules increase. This analysis shows that even for a truly ideal
system, the smallest simulation comprising only 20 molecules will generate noisy estimates
of the pair distribution function. This clearly demonstrates the problems inherent to small
systems where finite size effects are accentuated. Interestingly, for the ideal chain system, all
simulations with 100 or more molecules yield similar pair distribution functions as assessed
in Figure 3.3b.
Figure 3.3: Assessing finite size effects in simulations with ideal chains. (a) Pair distribution functions computed in terms of the spatial separation between chain units. The distributions are maximal at $r = L/2$, where $L$ is the size of the simulation cell for a given system. Note that $L$ increases as the number of polymers in the system increases. With the exception of the smaller systems, the ideal chains show self-similar behavior for different system sizes. (b) The data plotted in panel (a) are re-plotted in terms of the scaled variable $r^* = 2r \sqrt{3} L_i$ where $L_i$ is the size of the simulation cell for boxes with $i$ molecules. Note that all curves have some intrinsic roughness since the proteins live on a lattice.

Next we assessed the impact of finite size effects with all of the energetic terms included in the simulation. There are three columns, one each for different values of the reduced temperature $T^*$, in Figure 3.4. As discussed in the results section, these values of $T^*$ place the system of interest in the two-phase regime, with the quench depth into the two-phase regime increasing as $T^*$ increases. The first row (a to c) of Figure 3.4 shows the same data as Figure 3.3a while the second row (d to f) in Figure 3.4 show the normalized data like Figure 3.3b. Each panel shows eight unnormalized pair distribution functions, one each for the systems with 20, 100, 200, 1000, 1500, 2000, 3000, and 4000 molecules, respectively.
Figure 3.4: Assessing finite size effects in simulations with real chains. Panels (a), (b), and (c), respectively are the real chain equivalents of panel (a) in Figure 3.3 computed for three different simulation temperatures that represent three different quench depths of the system into its two-phase regime. Panels (d), (e), and (f), are the rescaled versions of panels (a), (b), and (c), respectively with the variable $r^* = \frac{2r}{\sqrt{3}L_i}$ where $L_i$ is the size of the simulation cell for boxes with $i$ molecules. Note that all curves have some intrinsic roughness since the proteins live on a lattice.

The onset of phase separation should be manifest by the presence of a trough located between two peaks in the profiles for $P^{(2)}(r^*)$. This is evident for $T^* = 0.267$ (Figure 3.4c) for all systems providing the numbers of molecules are greater than or equal to $10^3$. This qualitative trend is preserved even for $T^* = 0.217$, although sampling difficulties in large systems become obvious in the noisy estimates for $P^{(2)}(r^*)$. At the reduced temperature of $T^* = 0.167$ we confront two problems: The small systems where the numbers of molecules are less than $10^3$. 

89
cannot support the distinction between a proper dense phase that coexists with a dilute phase. This behavior is similar to that observed for lower quench depths i.e., higher simulation temperatures as shown in panels (b) and (c) of Figure 3.4. However, as the system size grows, an additional problem arises and this has to do with large clusters becoming frozen, and thus inhibiting the achievement of equilibrium. This is evident from the pair distribution functions shown in panels (a) and (d) of Figure 3.4 for systems where the numbers of molecules exceed $10^3$. To overcome this broken ergodicity and obtain reliable converged pair distribution functions, we need additional biasing potentials and temperature sweep approaches used recently [8] to break up frozen clusters and enable their coalescence. In the results that we report here, we use the system size titration to identify the reduced temperatures below which broken ergodicity becomes evident. We do not include data from these simulations in our constructions of phase diagrams. Importantly, our analysis confirms the presence of finite size effects for small systems and sets a lower bound on the numbers of molecules that are needed to observe facsimiles of phase separation as diagnosed by the calculated pair distribution functions. The conclusions drawn from analysis of the pair distribution functions are reinforced in our analysis of the radial distribution function shown in Figure 3.5.
Figure 3.5: The pair distributions from Figure 3.3 and Figure 3.4 are used to compute the relevant radial distribution functions. This analysis is relevant because the radial distribution functions are used to extract the value of the order parameter that detects the onset of phase separation. For systems where the number of molecules $A_n + B_n$ molecules is greater than 200, the radial distribution functions start to deviate from one another only at the lowest temperatures where broken ergodicity becomes an issue. Therefore, for the $A_n + B_n$ system studied in this calibration, it appears that the numbers of $A_n + B_n$ molecules have to be greater than 200 in order to obtain reliable information about the phase behavior. (a) extracted for different numbers of $A_n$ and $B_n$ molecules for $T^* = 0.167$. (b) extracted for different numbers of $A_n$ and $B_n$ molecules for $T^* = 0.217$. (c) extracted for different numbers of $A_n$ and $B_n$ molecules for $T^* = 0.267$. Note that unlike Figure 3.3 and Figure 3.4, the curves for $\tilde{g}(r^*)$ are smoother since the intrinsic roughness from the pair distributions cancels out.

### 3.3.4 Estimating The Percolation Transition Boundary

Associative polymers form networks characterized by physical crosslinks among stickers. Accordingly, we use the concept of a cluster, viz., a collection of unique chains connected via
anisotropic interactions, to define the extent of percolation. In polymer melt simulations, the extent of percolation, known as the gel fraction in the polymer literature, is defined as the fraction of polymers participating in a percolating network that spans the simulation box in at least one direction [9]. More generally, we can use the fraction of polymers that make up the single largest cluster to quantify the onset of percolation and the changes to the extent of networking beyond the percolation threshold [10]. A molecular network cannot percolate the whole simulation cell when dilute and dense phases coexist. Accordingly, we choose the second definition for the order parameter that describes the percolation transition, and we denote this as $\phi_c$ [11].

Semenov and Rubinstein demonstrated that a percolation transition is purely a connectivity transition [12]. This implies that the identification of the percolation threshold is not achievable using a bona fide order parameter but instead relies on a suitable topological description. Here, we employ the midpoint of the $\phi_c$ vs. concentration curve to assess the onset of percolation and the percolation line or curve is obtained as the locus of points in the phase diagram for which $\phi_c = 0.5$. In a system where finite size effects are minimized, the percolation transition is sharp having either a hyperbolic or sigmoidal shape as a function of concentration. Intuitively, note that for a system with $N$ molecules, there can only exist one cluster of size greater than $N/2$. Therefore, if the largest cluster is larger than half the system size, that cluster becomes unique. Accordingly, the location of the percolation line will be relatively robust to the choice one makes for the percolation threshold.
3.3.5 Results And Discussion

LaSSI Recapitulates FUS Phase Behavior

Figure 3.6 shows phase diagrams for the $A_n + B_n$ and $A_n - B_n$ systems calculated using data from LaSSI-based simulations. In panels (a) and (b), the ordinate quantifies the reduced temperature $T^*$ calculated as $T^* = k_B T / \epsilon$, where $\epsilon$ is the effective energy of pairwise interactions between stickers from the $A_n$ and $B_n$ blocks. Panel (a) shows results for the $A_n + B_n$ system. The bulk concentration in the $A_n + B_n$ system is quantified along the abscissa as $c_{\text{bulk}} = c_{A_n} + c_{B_n}$ where $c_{A_n}$ and $c_{B_n}$ are the bulk concentrations of $A_n$ and $B_n$, respectively. Panel (b) shows the phase diagram for the $A_n - B_n$ system where the abscissa represents the bulk concentration of this system.
Figure 3.6: Phase Behavior For A FUS-Like Linear Multivalent Protein. (a, b) Phase diagrams for the $A_n + B_n$ and $A_n - B_n$ systems, respectively. The purple line is a 2-dimensional linear interpolation for $\bar{\rho} = 0.025$, and the area encapsulated by the purple line are where the systems have large density inhomogeneities and are thus considered to be phase separated. The green line is a 2-dimensional linear interpolation for $\phi_c = 0.5$ and thus is the proxy for the percolation line. (c, d) and $\phi_c$ curves as a function of concentrations at $T^* = 0.383$ (solid lines in (a) and (b)). (e) Width of the two-phase regime, $w(T^*)$, as a function of the reduced temperature. Not only does the $A_n - B_n$ system have a higher critical temperature ($T^* \sim 0.6$ vs. $T^* \sim 0.4$), but also has a wider two-phase regime than the $A_n + B_n$ system.
Experiments show that the driving forces for phase separation are roughly twice as large for the full-length FUS compared to the system comprising of a 1:1 mixture of PLDs and RBDs [2]. This feature is recapitulated in LaSSI simulations. For example, the width of the two-phase regime is larger for the \( A_n - B_n \) system compared to the \( A_n + B_n \) system for all values of \( T^* \) as shown in panel (e) of Figure 3.6. The critical temperature is higher for the \( A_n - B_n \) vs. \( A_n + B_n \) system (\( T^* \approx 0.56 \) vs. \( T^* \approx 0.36 \), respectively). The valence of stickers is the main determinant of the concentration at the critical point whereas the interactions mediated by spacers determine the density inhomogeneities and the critical temperature. The impact of longer chains and increased valence of stickers per chain is also evident from the percolation threshold, which is crossed at a bulk protein concentration that is two-fold lower for the \( A_n - B_n \) system when compared to the \( A_n + B_n \) system, across all the simulation temperatures. Differences between the two systems are also evident in the degree of coupling between phase separation and the percolation transition as shown in panels (d) and (e) of Figure 3.6.

For each system, the intersection of the percolation threshold line with the two-phase regime shows that the dense phase predominantly forms a percolated droplet—panels (a) and (b) in Figure 3.6. Therefore, while phase separation without percolation is realizable, this is not the dominant scenario for associative polymers, where phase separation and percolation are typically conjoined to give rise to percolated droplets. The density of proteins in these percolated droplets is governed by the interaction strengths, modulated by \( T^* \) and the effective solvation volumes of spacers [11, 13]. Unlike homopolymers, which comprise entirely of stickers or spacers depending on the solvent quality, associative polymers encompass a mixture of stickers and spacers. Stickers provide the driving forces for networking via reversible crosslinks and spacers determine whether or not these driving forces lead to phase
separation via condensation. Indeed, the importance of sticker-driven percolation is evidenced in the persistence of percolated networks for both systems at high values of $T^*$. The observation that dense phases form percolated droplets has several implications: (1) on timescales that are concordant with or smaller than the average lifetime of physical crosslinks between stickers, the condensates will have elastic properties; this will be replaced by viscous behavior on timescales that are longer than the average lifetime of physical crosslinks [14]; (2) accordingly, condensates will have an intrinsic tendency for viscoelasticity [15] and long-lived crosslinks will cause hardening behavior as has been observed in many systems [2, 16–23]; (3) the extent of crosslinking above the percolation threshold will change continuously with concentration [11, 24], and this will govern the overall structure, internal dynamics, and porosity of condensates; (4) reactions within condensates are likely to be constrained by the network topology formed as a result of inter-sticker interactions [25]; these constraints can create a variety of interesting kinetic signatures for reactions [26], including temporal memories as has been demonstrated recently for a system that undergoes thermoresponsive phase behavior [27]. Clearly, any description of biochemical reactions within condensates has to account for the structural and dynamical attributes of percolated droplets that are best described as network fluids.

**Analyzing Move Acceptance Ratios**

Figure 3.7 shows the concentration dependence of acceptance ratios for each of the move set types, diagnosed for simulations of the $A_n + B_n$ and $A_n - B_n$ systems. The acceptance ratios show similar qualitative trends for both systems, even though there are clear quantitative differences. The move with the highest acceptance ratio in the dense regime is the double pivot move. The second most accepted move is the local move; extrapolating from the higher concentrations it is expected that the acceptance of local moves should also become small
and that the double pivot move will be the most dominant way to alter chain configurations, since even the move of an individual monomer will require that multiple monomers from multiple chains are moved simultaneously. Both systems have similar qualitative trends for the translation move where we see a transition from being accepted at low concentrations to not being accepted at higher concentrations. Since the proteins in the $A_n - B_n$ system are twice long as the $A_n + B_n$ system, the absolute acceptance ratio of the translation move is always lower in the $A_n - B_n$ system.

Figure 3.7: Analysis of move acceptance ratios for the $A_n + B_n$ and $A_n - B_n$ systems, respectively. The dashed lines the saturation concentrations. Data are obtained from simulations with $T^* = 0.383$. (a) Move acceptance ratio for the $A_n + B_n$ system. (b) Move acceptance ratio for the $A_n - B_n$ system.

Analysis of acceptance ratios of different move sets within droplets will be helpful for estimating correlation lengths and amplitudes of conformational and concentration fluctuations within droplets. Cluster moves have high acceptance ratios in the dilute regime whereas the
acceptance ratio nearly vanishes as the concentration increases. This is intuitive since the likelihood of steric clashes increases with a decrease in available volume and this is coupled to the simultaneous increase in the fraction of molecules in the largest cluster. We note here that the cluster moves have the most dramatic change in acceptance ratios from values near 1 to values near 0. However, the apparent inefficiency of cluster moves in dense configurations cannot be used as a rationale to quench such moves. In fact, as shown in panel (a) of Figure 3.8, phase separation, diagnosed in terms of $\bar{\rho}$, is suppressed if cluster moves are not part of the move set. This highlights the importance of cluster moves for generating bona fide phase separation as these facilitate coalescence that leads to condensation.

![Graphs showing phase separation](image)

**Figure 3.8:** (a) $\bar{\rho}$ and (b) $\phi_c$ curves for the $A_n + B_n$ (purple) and $A_n - B_n$ systems (green) at $T^* = 0.383$. The solid lines are identical with the curves in panels (c) and (d) of Figure 3.6. The dotted lines show the simulation results under the same system conditions but the frequency for cluster translation moves is set to zero. Not only do the systems phase separate and percolate at higher saturation concentrations, but also we can see that both percolation and separation are suppressed highly. Furthermore, errors are generally higher, due to the systems being highly dependent on the initial conditions of the system.
3.4 Branched Multivalent Proteins

In addition to demonstrating the applicability of the LaSSI engine for simulations of branched systems, we use the analyses as an opportunity to highlight key conceptual features of multicomponent systems that undergo phase separation via obligate heterotypic interactions. There are three main features that are borne out in the analyses: (1) For fixed temperature, the order parameters are the concentrations of the proteins that drive phase separation via obligate heterotypic interactions. In such systems, the coexistence curves will have a closed loop form. These can be ellipses for two components and $n$-ellipsoids for systems that involve up to $n$ obligate heterotypic interactions to drive phase separation. (2) The systems will support re-entrant phase behavior as has been reported recently for protein-RNA mixtures that undergo phase separation via obligate heterotypic interactions [28]. (3) The apparent saturation concentration of a component molecule in a system that undergoes phase separation via obligate heterotypic interactions will show non-trivial dependencies on its bulk concentration. These dependencies are governed directly by the slopes of the tie lines that pass through the point corresponding to the bulk concentration and intersect the binodal at coexisting concentrations that equalize the chemical potentials of all species in the dense and dilute phases. Here, we use the example of the N130 + rpL5 system to showcase the three central features of phase diagrams for systems that undergo phase separation via obligate heterotypic interactions.

3.4.1 N130 & rpL5 As An Archetypal System

As stated above, LaSSI can also be deployed to study the phase behavior of branched multivalent proteins that undergo phase separation via obligate heterotypic interactions. Examples of branched multivalent proteins are molecules that form symmetric, stable oligomers
such as nucleophosmin 1 (NPM1) and synthetic systems such as the corelets designed by Bracha et al. [29]. NPM1 is a key molecule within the granular component (GC) of the nucleolus [30]. Three coexisting layers define the nucleolus and the GC is the outermost layer. Within the GC, NPM1 appears to form facsimiles of percolated droplets in complex with ribosomal proteins with Arg-rich motifs [31, 32]. A minimalist system that mimics the phase behavior of the GC comprises of a truncated version of NPM1, referred to as N130, and an Arg-rich peptide, designated as rpL5 [3, 4, 33]. Both NPM1 and N130 form symmetric pentamers in the presence of Arg-rich peptides [34]. The pentamer formed by the association of folded domains serves as a scaffold for displaying disordered C-terminal tails that are defined by two distinct acidic tracts. The system also features an N-terminal disordered region with a well-defined acidic motif.

In the LaSSI representation, N130 pentamers with disordered tails are modeled using a five-armed structure. This approach follows the strategy of Feric et al. [32], which was based on the fact that pentamers do not dissociate under conditions where NPM1 / N130 undergo phase separation. Each arm comprises two sticker sites to mimic the presence of the A1 and A2 acidic tracts within the disordered tails of NPM1 / N130. Therefore, each N130 pentamer displays a total of ten sticker sites. The spacers between each A1 tract and the N130 core as well as between each pair of A1 and A2 tracts on a disordered tail are phantom tethers, which means that their effective solvation volumes [11] are set to zero. Each rpL5 peptide has two sticker sites corresponding to the two Arg-rich motifs along the sequence. Schematic representations of the coarse-grained architecture used for N130 and rpL5 are shown in Figure 3.9, and the move set frequencies are summarized in Table 3.2.
Figure 3.9: Coarse-grained representation of an archetypal branched multivalent system. (a) Cartoon schematic to depict the overall structure of the two molecules: N130; rpL5. The N130 pentamer has 5 arms where each arm has 2 stickers (orange beads) representing the acidic tracts – giving a total of 10 stickers per molecule. The central monomer (gray bead) is a spacer which represents oligomerization domain. The rpL5 peptide is modeled as having 2 stickers (blue beads) representing the basic motifs. (b) The relevant length-scales for the two molecules. The rpL5 peptide has a linker-length of 3 owing to its disordered nature, while for N130 the first sticker (modeling the A1 tract) is adjacent to the spacer bead with a linker length of 1, and the second sticker (modeling the A2-tract) is 3 lattice-sites from the first sticker.

3.4.2 Results And Discussion

Percolation Lines Have Parabolic Shapes

The percolation line, constructed as a function of the concentrations of two multivalent molecules, has an overall parabolic shape (panel (a) of Figure 3.10). This feature may be explained as follows: Let $f_1$ and $f_2$ denote the fractions of N130 and rpL5 molecules that are bound in a network; $s_1$ and $s_1$ will denote the valence of stickers on N130 and rpL5, respectively; for the current implementation of the N130 + rpL5 system, $s_1 = 10$ and $s_1 = 2$. Based on the Flory-Stockmayer theory [35, 36], the percolation threshold is crossed when
\( f_1 f_2 (s_1 - 1)(s_2 - 1) > 1 \). If we keep \((s_1 - 1)(s_1 - 1)\) constant, the threshold concentration for percolation will be governed by the product \(f_1 f_2\). Accordingly, if there is a large excess of N130 (component 1) compared to rpL5 (component 2), then \(f_1 \to 0\) and \(f_2 \to 1\), and the system does not undergo a percolation transition. In this scenario, every rpL5 molecule is crosslinked to two sticker sites from one or two N130 molecules. However, since the relative stoichiometry favors N130 molecules, there is a vast excess of un-crosslinked N130 molecules and the network cannot grow. Percolation is also inhibited when the converse situation arises, i.e., when there is a large excess of component 2 with respect to component 1. Accordingly, the percolation line takes on a parabolic form in the plane defined by the concentrations of N130 and rpL5.
Figure 3.10: Phase Behavior Of The Branched Multivalent Systems For $T^* = 0.25$ (a) The full phase diagram, where the purple line denotes the proxy for the binodal and the green line is the proxy for the percolation line (see also the caption for Figure 3.6). The phase-separated region has an elliptical shape, and we have a closed loop which represents re-entrant phase behavior. The percolation line, in contrast, has a conical shape extending into much higher densities. The solid black lines denote contours of constant total concentration where L1 is the lowest concentration, and L3 is the highest concentration. Note that that both axes are log-scale. (b, c) $\bar{\rho}$ and $\phi_c$ curves as a function of the relative stoichiometric ratio of rpL5 to N130 along the constant-concentration contours L1, L2, and L3. (d) Plot of $\Lambda$ as a function of the relative stoichiometry along the contours L1, L2, and L3.
Coexistence Curves For Systems With Obligate Heterotypic Interactions Form Closed Loops

Given that the phase behavior of the N130 + rpL5 system is driven by heterotypic interactions involving the A1 / A2 tracts from the N130 tails and the Arg-motifs from rpL5, we constructed coexistence curves by keeping the simulation temperature fixed and varied the concentrations of N130 and rpL5 molecules. Phase diagrams defined by N130 concentration along the abscissa and rpL5 concentration along the ordinate are shown in panel (a) of Figure 3.10. The general shape of the coexistence curve is comparable with that of the experimentally determined phase diagram [4], even though direct comparison is not straightforward because the scarcity of experimental data points does not yield a full coexistence curve.

The phase boundary, defined by the density transition, is an ellipse that forms a closed loop in the plane defined by the concentrations \( c_1 \) and \( c_2 \) of N130 and rpL5, respectively. In associative polymers, the phase behavior is governed by the affinity between stickers, the valence of stickers, and the effective solvation volumes of spacers [11, 13]. For fixed \( c_1 \) that intersects the two-phase regime an increase in \( c_2 \) will lead to an entry into the two-phase regime caused by a density transition as \( c_2 \) approaches \( c_1 \). However, as \( c_2 \) increases well beyond \( c_1 \), the joint system exits the two-phase regime. This is because phase separation is driven by obligate heterotypic interactions and while there is a growing excess of rpL5 molecules there are not enough N130 molecules to drive the density transition via inter-sticker interactions. Similar reasoning applies to describe the re-entrant behavior that will result by keeping \( c_2 \) fixed at a value that intersects the two-phase regime and increasing \( c_1 \).

Taken together, the parabolic percolation lines and elliptic forms for two-phase regimes define conic sections that highlight re-entrant phase behavior whereby fixing the concentration of component 1 and increasing the concentration of the second species can lead to phase
separation and percolation at a low threshold concentration of component 2 and exit into
the one-phase, non-percolated regime beyond a second higher threshold concentration for
component 2. This type of re-entrant phase behavior, will be a general feature of multicom-
ponent systems that undergo phase separation via obligate heterotypic interactions; indeed,
re-entrant phase behavior has been reported for a model protein + RNA system [28].

**Apparent Stoichiometric Ratios Can Differ From Effective Stoichiometric Ratios**

Stoichiometry of molecules that drive phase separation is another key parameter that deter-
nines the functions of biomolecular condensates formed by multicomponent systems [37].
The apparent stoichiometric ratio is calculated as the ratio of the concentrations of stickers of
types $s_2$ and $s_2$ for N130 and rpL5, respectively, such that $\nu_{12}^{\text{app}} = \frac{c_{s_1}}{c_{s_2}}$. However, the effective
stoichiometric ratio $\nu_{12}^{\text{eff}}$ can be different from $\nu_{12}^{\text{app}}$ if excluded volume effects modulate the
effective concentration of stickers. We fit an ellipse to the two-phase boundary and determined
the major axis of this ellipse. The effective stoichiometric ratio should be unity along the
major axis. As shown in panel (a) of Figure 3.10, the major axis deviates from the line along
which $\nu_{12}^{\text{eff}} = 1$. Therefore, $\nu_{12}^{\text{app}} \neq \nu_{12}^{\text{eff}}$ and angle between the major axis and the line along
which $\nu_{12}^{\text{eff}} = 1$ quantifies the impact of excluded volume on changes to effective concentrations
of stickers that in turn modifies the stoichiometric ratios.

The synergy between stoichiometry and phase behavior can be analyzed by quantifying the
order parameter and the topological parameter $\phi_c$ as a function of apparent stoichiometry
for fixed bulk concentration. Along each gray line in panel (a) of Figure 3.10 the total
concentration defined as $c_{\text{bulk}} = (c_1 + c_2)$ is fixed, although the stoichiometries will vary. The
mean values of $c_{\text{bulk}}$ along L1, L2, and L3 are $2.09 \times 10^{-2}$ (voxel$^{-1}$), $2.46 \times 10^{-3}$ (voxel$^{-1}$), and
$3.33 \times 10^{-4}$ (voxel$^{-1}$), respectively and the value of ranges from 0.36 to 22.62 along each of
L1, L2, and L3. Panels (b) and (c) in Figure 3.10 show the variation of $\bar{\rho}$ and $\phi_c$ as increases
along L1, L2, and L3, respectively. Along L1, the value of $\rho$ is essentially zero irrespective of stoichiometry because L1 lies is outside the two-phase regime. However, a system spanning percolated network forms for stoichiometries in the range $1.2 \leq \nu_{12} \leq 13$ along L1. This is because the concentrations of both components are well above the percolation threshold along L1 thus ensuring that stickers readily find one another even without a density transition. In direct contrast, along L3, we observe phase separation, characterized by values of $\rho > 0.025$ for a range of stoichiometries, but none of these support the formation of a percolated droplet ($\phi_c < 0.5$ for all stoichiometries). Along L2, we observe phase separation for stoichiometries in the range $1.15 \leq \nu_{12} \leq 16$ and percolation for stoichiometries in the range $2.14 \leq \nu_{12} \leq 11.3$ such that phase separation enables the formation of a percolated droplet.

In panel (d) of Figure 3.10, we introduce a new structural parameter $\Lambda$, which we define as a convolution of $\rho$ and $\phi_c$ such that $\Lambda = \rho \times \phi_c$. Here, the convolution is calculated as a logical AND gate, which becomes a simple product. The parameter $\Lambda$ quantifies the convolution of the density and network transition and provides an estimate of the extent to which the phase separation and percolation are coupled as the apparent stoichiometry is varied for a fixed bulk concentration. The profile of $\Lambda$ is reminiscent of profiles measured by Case et al. [37] for the dwell time of signaling molecules as a function of stoichiometric ratios that govern the formation of condensates at membranes. This suggests that dwell times, which are experimentally accessible parameters, might actually be proxies for the structural features of the condensates as measured by the convolution between phase separation and percolation and the extent of network formation within the condensate.

The key finding is that the combination of the bulk concentration and stoichiometric ratio (as opposed to stoichiometry alone) will determine the quench depth into the two-phase regime. This in turn determines whether a system-spanning network forms without phase separation or if phase separation enables the formation of a droplet-spanning network. The structure of
condensates and the overall phase behavior cannot be fully described in terms of $c_{\text{bulk}}$ or $\nu_{12}$ alone. Instead, this requires the consideration of both parameters jointly and relative to the quench depth, which refers to the location in the two-phase regime and with respect to the percolation line. This is important because the extent of crosslinking and the time scales associated with crosslinks will determine the material properties of the condensate. This in turn should contribute to parameters such as the dwell times of clients within condensates [37].

The Concept Of Saturation Concentration Breaks Down In Multicomponent Systems

The concept of a saturation concentration is one of the defining hallmarks of phase separation [38, 39]. For fixed solution conditions, phase separation in a closed two-component system (or pseudo one-component system) comprising of a protein and solvent is realized when the bulk concentration of the protein denoted as $c$ exceeds a saturation concentration denoted as $c_{\text{sat}}$. The presence of a saturation concentration can be quantified by measuring the concentration of protein in the coexisting dilute phase as $c$ increases. This value will not go above $c_{\text{sat}}$. Strikingly, the presence of a saturation concentration has been confirmed in living cells for model disordered proteins by Brangwynne and coworkers using optogenetic tools in mammalian cells [29, 38] and by Khan et al. [40] using yeast as a model system.

If the protein whose phase behavior is being interrogated is part of a system where obligate heterotypic interactions drive phase separation, then whether or not the concept of a saturation concentration continues to be valid will depend on the nature of the binodals. We illustrate this by coopting the elliptic phase boundary from Figure 3.10 for a three-component system that comprises of N130, (component A), rpL5 (component B), plus a solvent that is implicit
in the LaSSI simulations. This 3-component system may be thought of as a pseudo two-component system. For fixed temperature, the top row of panels (a)-(c) in Figure 3.11 show three types of elliptical, closed loop binodals. These are constructed in a plane where [B] increases along the positive direction of the abscissa and [A] increases along the positive direction of the ordinate. The bottom row in each panel shows the result to be expected were we to measure the concentration of A in the dilute phase, designated as $[A]_{\text{Sol}}$, as the bulk concentration [A] is varied. In each of these measurements, the concentration of B is fixed at a specific value.
Figure 3.11: Slopes of tie-lines with elliptical binodals are important for systems that undergo phase separation. The ellipse is drawn to fit the locus of points such that $\bar{\rho} = 0.025$, representing the boundary for a density transition (main text, Sub-section 3.3.2) for a two-component system where macromolecule A and B undergo obligate heterotypic interactions. The data to construct the ellipse were taken from simulations of the N130 + rpL5 system shown in Figure 3.10. For the same ellipse, and thus phase boundary, we draw hypothetical tie-lines of different slopes in each column to assess how the dilute phase for A changes. The three lines colored green, gray, and blue, respectively represent different total B concentrations where we vary the concentration of A. (a, b) Ellipse annotated with nearly horizontal tie-lines. Despite the system undergoing a phase transition, the curves for $[A]_{sol}$ do not show canonical saturation behavior where $[A]_{sol}$ remains fixed after a particular [A]. (c, d) Ellipse annotated with upwards diagonal tie-lines, closer to the N130 + rpL5 system. Again, we do not see the canonical saturation behavior for $[A]_{sol}$; but instead we see that $[A]_{sol}$ decreases as we enter the two-phase regime. This is because the dilute phase now goes along the lower boundary of the ellipse. (e, f) Ellipse annotated with nearly vertical tie-lines. Finally, we see canonical saturation behavior. For a given [B], as we increase [A] we get plateauing of $[A]_{sol}$ inside the two-phase regime. The plateau abruptly ends as we exit the two-phase regime.
For concentrations of A and B that place the pseudo two-component system in the two-phase regime–red points along each of the curves in the bottom rows of panels (a)–(c)–we find that $[A]_{\text{Sol}}$ can change as $[A]$ increases. If the tie lines are horizontal or nearly horizontal, then $[A]_{\text{Sol}}$ will vary linearly with $[A]$. Non-linear variations of $[A]_{\text{Sol}}$ with $[A]$ will result for tie lines with positive or negative slopes. This is shown in panel (b) for tie lines with positive slopes. If the tie lines are essentially vertical, then the standard expectation regarding the invariance of $[A]_{\text{Sol}}$ with $[A]$ within the two-phase regime is recovered. However, even in this scenario, the plateau value of $[A]_{\text{Sol}}$ will shift upward or downward as the value of $[B]$ increases—the upward shift is shown in panel (c) of Figure 3.11. Here, B acts as a \textit{bona fide} ligand for A, which is the macromolecule. Preferential binding of B to the dilute phase leads to an increase in $[A]_{\text{Sol}}$ as depicted in panel (c) of Figure 3.11. Ligand-mediated shifts in saturation concentrations arise due to polyphasic linkage, a phenomenon first introduced by Wyman and Gill [41]. This ligand-mediated modulation of phase behavior is further explored in Chapter 4.

3.5 Conclusions

In this work, we have built on the connection between multivalent proteins and associative polymers [12, 25, 42] with their stickers-and-spacers architecture [2, 8, 11, 13, 22, 31, 43, 44] to develop and deploy LaSSI, a lattice-based open source computational engine that enables the simulation of system-specific phase diagrams of single and multi-component systems. The foundations of LaSSI derive from the formalism of the bond fluctuation model [45–47]. We demonstrate how canonical ensemble Monte Carlo simulations with appropriately designed move sets and analysis of order parameters derived from the distribution functions allow us to calculate coexistence curves and percolation lines as a function of protein concentration and interaction strengths.
The choice of a lattice-based approach for coarse-graining and modeling phase behavior of multivalent proteins is guided by the advantages of lattice models [48] for polymeric systems. To titrate across the full range of volume fractions, one needs to balance considerations of finite size effects—which requires large numbers of molecules—with large simulation volumes—which makes it difficult to observe density fluctuations that can grow into density inhomogeneities. On lattices the conformational space is discretized and the calculation of interaction potentials can be made to be very efficient through the use of look up tables. Importantly, we have generalized lattice-based simulations to incorporate anisotropic interactions.

LaSSI allows us to query the impacts of overall and intrinsic valence of stickers, interaction strengths between stickers, the spatial ranges of these interactions, the effective solvation volumes and lengths of spacers, and protein concentrations. These titrations generate multidimensional phase diagrams. The approaches underlying LaSSI have been applied to model a variety of multicomponent systems, including mimics of RNA molecules [2, 8, 11, 13, 32, 49]. What is required is the development of approaches that enable systematic coarse-graining and adaptation of machine learning based methods to parameterize interaction potentials [50]. Engineering LaSSI to be interoperable to cell-based modeling suites [51] will also allow for larger scale deployment of the overall framework. The calculation of pair and higher order distribution functions should afford multiscale descriptions of the structural organization of molecular components within condensates. The acceptance ratios associated with different move sets and the length scales spanned by distinct move sets open the door to analyzing the dynamics of phase separation, percolation, and the interplay between the two. Another major direction for future application of LaSSI is to uncover the determinants of compositional specificity of condensates [16, 52].
3.6 References


7. Deserno, M. How to calculate a three-dimensional g(r) under periodic boundary conditions (2004).


    Bickle, M., Rizk, S., Guill én-Boixet, J., Franzmann, T. M., Jahnel, M., Marrone, L.,
    Chang, Y.-T., Sterneckert, J., Tomancak, P., Hyman, A. A. & Alberti, S. RNA buffers
    the phase separation behavior of prion-like RNA binding proteins. *Science* 360,

    J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T. & Rosen,
    M. K. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483,


27. Roberts, S., Harmon, T. S., Schaal, J. L., Miao, V., Li, K., Hunt, A., Wen, Y., Oas,
    T. G., Collier, J. H., Pappu, R. V. & Chilkoti, A. Injectable tissue integrating networks
    from recombinant polypeptides with tunable order. *Nature Materials* 17, 1154–1163
    (2018).

    Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets.

    J. E. & Brangwynne, C. P. Mapping Local and Global Liquid Phase Behavior in Living


Chapter 4

Ligand Effects On The Phase Separation of Multivalent Macromolecules

4.1 Preamble

LASSI-based simulations. K.M.R. prepared all the figures. R.V.P. secured funding. All authors read and contributed edits to the final version of the manuscript.

4.2 Introduction

Membraneless biomolecular condensates concentrate biomolecules in cells to organize biochemical reactions in space and time [3, 4]. There is growing evidence that condensates form via spontaneous or driven phase transitions [4, 5]. Functional condensates can be reconstituted in vitro and manipulated in live cells using only one or a small category of macromolecular scaffolds [6–15]. Multivalence of interaction motifs known as stickers is a defining hallmark of scaffolds that drive phase transitions [10, 11] by combining density transitions in the form of phase separation and networking transitions in the form of percolation [6, 16, 17].

Mutations to scaffold molecules are associated with disease, and evidence is growing that these lead to changes in scaffold phase behavior [18]. Changes include lowering the threshold scaffold concentration needed for phase separation and lowering the barrier for liquid-to-solid transitions within condensates as summarized in the Table A.1. These results suggest that the formation and dissolution of condensates has to be tightly regulated in cells. One route to modulating phase behavior is through posttranslational modifications to scaffold molecules [19].

A second mechanism takes advantage of the fact that condensates contain several types of nonscaffold molecules. The expression levels of these nonscaffold molecules can be used as knobs that can be turned to control scaffold phase behavior (Table A.2 and Figure 4.1). We define these modulatory nonscaffold molecules as ligands.
Ligands do not undergo phase separation on their own and are not required for the phase separation of scaffolds. However, they bind preferentially to scaffolds across phase boundaries and either promote or destabilize condensates in cells. This is important because in some situations, the cellular concentrations of scaffolds may be too low to drive phase separation [20]. This can be remedied by the controlled expression of specific ligands that lower the concentration threshold for phase separation (Figure 4.1A). Alternatively, for scaffolds that can phase separate on their own at endogenous concentrations, increasing the expression level of a destabilizing ligand can help dissolve the condensate [21] (Figure 4.1A). Similar ideas can be brought to bear in designing pharmaceutical approaches to regulate condensates (Figure 4.1B).

![Figure 4.1: Schematic showing how ligands can modulate phase behavior in-vivo. (A) Examples of how tuning the expression levels of ligands can be used to regulate scaffold phase behavior under normal cellular conditions. (B) Example of how a ligand can be used to revert the phase behavior of a mutant scaffold back to wild-type phase behavior. Here, ligand expression levels are tuned from zero to high expression levels and green circles imply condensate formation of a fluorescently tagged scaffold. Hence, the brightness denotes the relative concentration of the scaffold in each phase.](image-url)
To understand the mechanisms that underlie the regulation of scaffold phase behavior by ligands, we coopt the polyphasic linkage formalism of Wyman and Gill [22]. Although this formalism was introduced four decades ago, it has not been deployed to understand, interpret, or appreciate the true scope of ligand-modulated phase separation in vitro and in live cells. Here, we establish why polyphasic linkage is useful for understanding how ligands can alter scaffold phase behavior. This, as explained above, is of direct relevance for understanding biological and pharmaceutical regulation of condensates in vivo (Figure 4.1).

To illustrate the concepts of polyphasic linkage, we consider an aqueous solution with a single type of scaffold that separates into two distinct phases. We denote the scaffold-deficient phase as A and the coexisting scaffold-rich phase as B. The binodal or coexistence curve delineates the two-phase regime (Figure 4.2). For a given set of solution conditions, quantified in terms of an effective interaction strength, the left arm of the phase boundary (the binodal) denotes the saturation concentration $c_A$ in the dilute phase, and the right arm of the binodal corresponds to the concentration $c_A$ of the scaffold in B, which is the dense phase. The polyphasic linkage formalism describes how ligand binding modulates $c_A$ for fixed solution conditions. The value of $c_A$ in the presence of the ligand, designated as $c_A^L$, is determined by equalizing the chemical potential of the scaffold across the phase boundary. This yields the following expression: $c_A^L = c_A \left( \frac{P_A}{P_B} \right)$ [22].
Figure 4.2: Schematic showing how ligands can modify scaffold binodals. Scaffolds undergo phase separation and form two distinct phases: the scaffold-deficient phase (red), $A$, and the scaffold-rich phase (blue), $B$. The yellow region delineates the two-phase regime of the scaffold. The phase diagram is drawn in the two parameter space of total scaffold concentration, $[S_T]$, and effective interaction strength among scaffold molecules. For a given value of the interaction strength along the ordinate, the left arm of the binodal represents the saturation concentration, $c_A$, and the right arm of the binodal denotes the concentration of the scaffold in the dense phase, $c_B$ (grey circles). Polyphasic linkage describes how $c_A$ changes by way of preferential binding of the ligand across the phase boundary and this is denoted as $c_A^L$. (A) $c_A^L > c_A$ if the ligand preferentially binds to the scaffold in the dilute phase. (B) $c_A^L < c_A$ if the ligand preferentially binds to the scaffold in the dense phase.
Here, $P_A$ and $P_B$ are the binding polynomials that quantify binding of the ligand to the scaffold in phases A and B, respectively. If $P_A$ is greater than $P_B$, then the ligand binds preferentially to the scaffold in phase A. Preferential binding of the ligand to the scaffold in the A phase will lead to an increase in $c_A^L$ compared to $c_A$, thus weakening the driving forces for phase separation of the scaffold (Figure 4.2A). Conversely, if $P_B$ is greater than $P_A$, then the ligand binds preferentially to the scaffold in the dense phase B. Accordingly, $c_A^L$ decreases when compared to $c_A$. In this scenario, preferential binding of the ligand to the scaffold in the dense phase will enhance the driving forces for phase separation as evidenced by lowering of the threshold concentration to be crossed in order for the system to undergo phase separation (Figure 4.2B). In the absence of preferential binding, the ligand binds equivalently to the scaffold in both phases, implying that $c_A^L = c_A$ and the ligand behaves like a passive client that does not alter the phase equilibrium of the scaffold.

A weakness of the polyphasic linkage formalism is that it does not yield information regarding the specific features of ligands that lead to preferential binding to one phase over the other. Here, we remedy this using the stickers-and-spacers model [16, 23] to uncover rules for how precise control over scaffold phase behavior can be achieved through the preferential interactions of ligands. To uncover these rules, we use a coarse-grained linear polymer model that mimics well-known examples of scaffolds [3, 6, 10–12, 15, 24–26]. We assess the effects of five types of ligands. These include two that were previously examined [27–29] and three new ligand types, namely monovalent ligands that interact with scaffold stickers, monovalent ligands that interact with scaffold spacers, and bipartite ligands that interact with scaffold stickers and spacers. Through the use of these models, we generate insights regarding the features of ligands that stabilize or destabilize condensate formation by scaffolds. Importantly, we determine how specific features of ligands contribute to modulating scaffold phase behavior, provide mechanistic explanations for why ligands promote or destabilize
scaffold phase behavior, and quantify how the structure and concentration of the dense phase change upon ligand binding. Our work builds on key contributions reported recently that use patchy, spherical particles as models for scaffolds and ligands [27–29]. These studies have helped elucidate certain aspects of how low-valence patchy ligands exert control over the phase behavior of high-valence patchy scaffolds [29].

In addition to generating insights regarding ligand-mediated regulation of condensates, we highlight the importance of directly measuring how ligands affect phase boundaries of scaffolds. Such measurements, performed in live cells, are likely to pave the way for understanding how cells control condensate formation and dissolution via preferential binding of specific types of ligands at the right place and at the right time. Our analysis also shows that uncovering the modulatory effects of ligands cannot be achieved by measuring partition coefficients (PCs) of ligands. This is because the PC of a ligand is a convolution of many factors. Accordingly, high values of PCs for ligands do not have to mean a preferential binding to scaffolds in the dense phase, nor do they tell us anything about the modulatory effects of ligands on scaffold phase behavior.

Lastly, many cellular condensates are driven by a combination of homotypic and heterotypic interactions. For instance, in stress granule formation G3BP1 interacts with itself through its NTF2L dimerization domain and engages in heterotypic interactions with nuRNA molecules. Although, heterotypic interactions with RNA are dominant determinants of phase separation, the dimerization domain of G3BP1 is necessary although insufficient for phase separation [8, 9, 15]. Additionally, some of the putative scaffolds of P-bodies engage in homotypic interactions, including Edc3 [30–33]. Homotypic and heterotypic interactions of NPM1 may also be important for the formation of the granular component of the nucleolus [34, 35]. Therefore, as a demonstration of the need to further generalize ligand mediated changes to phase behavior in systems with multiple components, we study the effects of ligands for a
two-component system where the interactions are heterotypic only, or a mixture of both heterotypic and homotypic.

4.3 Materials And Methods

4.3.1 Coarse-Grained Model

We deployed coarse-grained simulations using the lattice simulation engine LaSSI [23] to understand how different types of ligands modulate the phase behavior of model linear multivalent macromolecules (details in Sub-section 4.3.2). Linear multivalent macromolecules are represented by stickers, which are sites that drive phase separation, and spacers, which are sites interspersed between stickers that influence the interplay between phase separation and percolation [23, 24]. Here, spacers can be implicit, in that they do not take up volume, or they can be explicit, in that they occupy volume on the lattice.

Previous studies have examined how the valence, sticker interaction strengths, and spacer excluded volumes of linear multivalent macromolecules effects scaffold phase behavior [6, 17, 24, 25]. Accordingly, we focus here on a single type of scaffold molecule. In our model system, the scaffold molecule contains five sticker sites and two explicit spacer sites (Figure 4.3A). The inclusion of explicit spacer sites allows for modeling the effects of ligands that interact directly with spacer sites. The specific instantiation of the stickers-and-spacers model used here helps ensure that: (a) the valence of spacer sites is less than the valence of sticker sites such that phase separation is driven mainly by sticker-sticker interactions regardless of ligand type; and (b) we observe robust phase separation in the concentration regime examined – an important design criterion given that increasing the number of explicit spacer sites can destabilize phase separation [24].
To determine the effect of ligand type on scaffold phase behavior, we consider five different ligand types (Figure 4.3C): (i) a monovalent ligand that interacts exclusively with scaffold stickers; (ii) a monovalent ligand that interacts exclusively with scaffold spacers; (iii) a divalent ligand that interacts exclusively with scaffold stickers; (iv) a divalent ligand that interacts exclusively with scaffold spacers; and (v) a bipartite divalent ligand that interacts with scaffold stickers and spacers. The only attractive interactions in the system are between pairs of sticker sites on scaffolds and between scaffold sticker or spacer sites and ligands. The details of the latter will depend on the type of ligand being considered. Additionally, each site can engage in only one interaction at a time. We consider three energy scales for the ligand-scaffold interaction (Figure 4.3B): $E_1$, the ligand-scaffold interaction is half that of the scaffold-scaffold interaction, $E_2$, the ligand-scaffold interaction is equal to the scaffold-scaffold interaction, and $E_3$, the ligand-scaffold interaction is double that of the scaffold-scaffold interaction. For each case, we performed five independent simulations as detailed in Sub-section 4.3.2.

### 4.3.2 Simulation Details

LaSSI [23] is designed for simulations of phase transitions of polymers on a lattice using Monte Carlo (MC) simulations. In the current simulations, 2000 chains of the macromolecular scaffold were used for every solution condition. To only have inter-molecular interactions, the scaffold was divided into two sets comprised of 1000 chains where each set had different sticker and spacer types. The total set of ligand numbers sampled corresponds to 400, 800, ..., 4000 where for the monomer ligands the numbers were doubled to keep the ligand-site number and concentration consistent. Periodic boundary conditions were used, with a box size of $L = 103$ lattice units. Each bead can only form one physical bond where bonds are stochastically formed in a way that preserves detailed balance over the course of the
simulations. Mathematical details regarding the sampling techniques used are as prescribed in the original work on LaSSI [23].

For any given solution condition (simulation temperature and combination of scaffold and ligand concentration), all molecules are randomly placed on the lattice and site-overlap is strictly forbidden. All interactions, other than site-overlap, are then turned off between the molecules. To accelerate the formation of one droplet for the dense phase, an external biasing potential is then applied to every bead which has the form $V(r) \sim (\vec{r} - \vec{r}_0)^2$ which serves as a means to push molecules towards the center of the simulation box. After $t_c = 5 \times 10^6$ MC sweeps, interactions between the molecules are turned on. The system is then simulated for $t_T = 2.5 \times 10^9$ MC sweeps for what we shall define as the first cycle.

Initially, the system is exponentially cooled down to $T_0 \equiv T^* = 1$, where the external potential now has the form $V(r, t) \sim e^{-\frac{4t}{t_c}}(\vec{r} - \vec{r}_0)^2$. When $T - T_0 \leq 0.005$, at $t_V \approx 7 \times 10^6$ we turn off the external potential ($V(r, t) = 0$). This results in the formation of only one large condensate provided the solution conditions correspond to the two-phase regime. The system is then run up till $t_T$. This concludes the first cycle.

The temperature of the system is then discontinuously increased by $\Delta T^* \approx 0.042$, and the system is simulated for $t_T$ MC sweeps, corresponding to the second cycle. 25 cycles are performed in total which sample $T^* = 1$ to $T^* = 2$.

5 independent sets of simulation are performed per solution condition using the following relative frequencies for the MC moves:
### Move frequencies

<table>
<thead>
<tr>
<th>Move</th>
<th>Normalized Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>3000</td>
</tr>
<tr>
<td>Bond Rotation</td>
<td>1500</td>
</tr>
<tr>
<td>Co-local</td>
<td>1500</td>
</tr>
<tr>
<td>Multi-local</td>
<td>1500</td>
</tr>
<tr>
<td>Chain Translation</td>
<td>500</td>
</tr>
<tr>
<td>Chain Pivot</td>
<td>500</td>
</tr>
<tr>
<td>Double-pivot</td>
<td>500</td>
</tr>
<tr>
<td>Limited Cluster</td>
<td>10</td>
</tr>
<tr>
<td>Large Cluster</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.1: Move frequencies for simulations. Note that the frequencies are normalized such that the least frequent move is scaled to 1.

Averages over the requisite quantities defined below are performed over the last half of each cycle. Lastly, note that statistically similar results can be obtained without the initial external potential, but the simulations need to be about 10 times longer. For more details about the MC engine in LaSSI, see [23], and Chapters 2 and 3.

### 4.3.3 Constructing Binodals

For every solution condition tested, we directly compute the co-existing densities via density-profiles referenced from the center-of-mass (COM) of the system. The density-profile is generated by computing a number histogram, $H(r_i)$, from the COM of the system with a bin-width of 0.25, up to $r = L\sqrt{3}/2$, which corresponds to the maximal distance in a cubic box with periodic boundaries. To normalize the number histograms, we explicitly
calculate the number histogram, $H_0(r_i)$, of lattice sites for a cubic box of dimensions $L = 10^3$ lattice-units with periodic boundaries, given a bin-width of 0.25. Thus, the density profile is

$$\rho(r_i) = \frac{H(r_i)}{H_0(r_i)}.$$  \hfill (4.1)

To calculate the dense phase concentration, $c_B$, we average over the first 7 bins excluding $r_0$, and for the dilute phase concentration, $c_A$, we average over the last 25 bins. For the binodals shown in Figures 4.3–4.6, only the co-existing densities are plotted where the densities are statistically different from each other.

### 4.3.4 Generating Site Molecularity Profiles

For each simulation, we compute the radial molecularity profiles using radial pair-number histograms. The radial pair-number histograms are generated using the standard procedure where, assuming isotropy, the radial singlet-number histogram, $H^{(1)}(r_i)$, is computed for a single bead. This radial histogram is then computed for every bead in the system where the sum over all beads then gives us the radial pair-number histogram $H^{(2)}(r_i)$ for the system. A bin-width of 1.0 is used, for smoothness, and the histogram is computed up to $r = L\sqrt{3}/2$. Since we can track the bead-type of every bead, we can compute this pair-number histogram for every possible type-pair. To normalize the distributions, we divide the histogram by the total number of unique possible pairs for a given set of types being considered. For each type-pair, the number of pairs is $N_{ij} = N_i(N_j - \delta_{ij})/(1 + \delta_{ij})$ where $N_{i,j}$ represents the number of beads of type $i, j$ and $\delta_{ij}$ is Kronecker delta function. In particular, we focus on two sets: sticker-sticker and (scaffold)-sticker-ligand.
Suppose that the sticker bead-types are 1 and 2, and that the ligand bead-types are 3 and 4. Then for the sticker-sticker set we have 1-2 pairs only, while for the second set we have the sum of 1-3, 1-4, 2-3 and 3-4 pairs. To compute the radial molecularity profiles from these normalized histograms, we multiply each histogram by the total number of beads, $\tilde{N}$, of the second index in the set. For the sticker-sticker set we multiply by the number of sticker beads of either type, $\tilde{N} = 5000$. For the (scaffold)-sticker-ligand set we multiply by the total number of ligand beads at a given concentration $\tilde{N}_m = 800m$ where $m$ is the concentration index. Therefore, the sum over each radial molecularity histogram is the number of sticker beads of either type for the first set and the number of ligand beads for the second set.

### 4.3.5 Constructing Binding Polynomials

Here, we go over a quick instructional derivation of binding polynomials for scaffolds. We shall assume that the scaffold ($S$) has $n$ independent ligand ($L$) binding sites, where the the ligand only has 1 binding site. The reaction equations for this process are given by:

$$[S] + i[L] \rightleftharpoons [SL_i]$$

where $i = 0, 1, 2, \ldots n$, $[SL_i]$ is the concentration of scaffold with $i$ ligands bound, and $[S]$ and $[L]$ are the free scaffold and ligand concentrations, respectively. Then, explicitly calculating the concentration of ligand-bound scaffold, we have

$$[SL] = [S_1L] + [S_2L] + \cdots + [S_nL],$$

131
where \([S_iL]\) represents the concentration of scaffold with site \(i\) bound. Following [22, 36], let \(k_j\) be the association constant for site \(j\). Then, the concentration of the scaffold bound to a single ligand is:

\[
[S_L] = k_1[S][L] + k_2[S][L] + \cdots + k_n[S][L]
\]  \hspace{1cm} (4.4)

where \([S_1L]\) denotes a single ligand bound to site 1 on the scaffold. Thus, the cumulative association constant for the scaffold bound to a single ligand is:

\[
\beta_1 = \sum_{j=1}^{n} k_j = \frac{[S_L]}{[S][L]}.
\]  \hspace{1cm} (4.5)

In general, the cumulative association constants are given by:

\[
\beta_i = \frac{[SL_i]}{[S][L]^i}.
\]  \hspace{1cm} (4.6)

The binding polynomial is the summation of the concentration of all states in the system involving the scaffold relative to the free scaffold,

\[
P = \sum_{i=0}^{n} \frac{[SL_i]}{[S]}.
\]  \hspace{1cm} (4.7)

Rewriting \(P\) in terms of cumulative associations constants, we get

\[
P = \sum_{i=0}^{n} \beta_i[L]^i.
\]  \hspace{1cm} (4.8)
Lastly, for an intuitive example, if all binding sites are independent and identical, \( k = k_1 = k_2 = \cdots = k_n \), and the cumulative association constant takes the form,

\[
\beta_i = \left( \frac{n!}{(n-i)!i!} \right)^{k^i}, \quad (4.9)
\]

[36], where we note that \( k^i \) is multiplied by the binomial coefficient \( \binom{n}{i} \). We can therefore rewrite Equation (4.8) as:

\[
P = (1 + k[L])^n. \quad (4.10)
\]

### 4.3.6 Calculating PC’s From Polyphasic Linkage Theory

With the general spirit of constructing binding polynomials within us, we shall now try to assess how PC’s of the ligand can be calculated from linkage theory. In particular, we shall consider an aqueous solution with a single type of scaffold that separates into two distinct phases. We denote the dilute phase as \( A \) and the coexisting dense phase as \( B \). For a system where ligand binding to the scaffold is described by a first order polynomial in both phases, ligand binding is given by:

\[
[S_A] + [L] \rightleftharpoons [S_AL] \quad (4.11)
\]

\[
[S_B] + [L] \rightleftharpoons [S_BL] \quad (4.12)
\]

133
Here, \([L], [S_A], [S_A], [S_B L]\) and \([S_B L]\) the concentrations of free ligand, free scaffold in phase A, free scaffold in phase B, bound scaffold in phase A, and bound scaffold in phase B, respectively. The binding polynomial in the dilute phase is given by:

\[
P_A = \frac{[S_A] + [S_A L]}{[S_A]} = 1 + k_A[L];
\] (4.13)

Here, \(k_A\) is the association constant in phase A. Likewise, the binding polynomial in the dense phase is \(P_B = 1 + k_B[L]\), where \(k_B\) is the association constant in the dense phase, \(B\). Accordingly, it follows that:

\[
c_A^L = c_A \left( \frac{P_A}{P_B} \right) = c_A \left( \frac{1 + k_A[L]}{1 + k_B[L]} \right).
\] (4.14)

It should be noted that this equation may not be valid for high concentrations of free ligand because it ignores non-idealities due to interactions of ligands with themselves. Additionally, it assumes that \(c_B\) does not change upon addition of the ligand.

To assess how \(c_A^L\) and \(PC\) are related we must solve the system specific set of equations that describe each binding reaction. Here, the system is defined by the following set of equations:

\[
k_A = \frac{[S_A L]}{[S_A][L]}, \quad k_B = \frac{[S_B L]}{[S_B][L]},
\] (4.15)

\[
[S_{A,T}] = [S_A] + [S_A L]; \quad [S_{B,T}] = [S_B] + [S_B L],
\] (4.16)
\[ [L_T] = [L] + [S_A L] + [S_B L]. \]  

(4.17)

Here, \([S_{A,T}]\) and \([S_{B,T}]\) are the total scaffold concentrations in phases \(A\) and \(B\), respectively, that are calculated using the total system volume \(V\). Likewise, \([L_T]\) is the total ligand concentration in the system. The total scaffold concentration is given by:

\[ [S_T] = \frac{n_A + n_B}{V_A + V_B} = \frac{n_A + n_B}{V} = [S_{A,T}] + [S_{B,T}]. \]  

(4.18)

Here, \(n_A, n_B\) are the numbers of scaffold molecules within and \(V_A, V_B\) are the volumes of phases \(A\) and \(B\), respectively; \(V\) is the total volume of the system. Accordingly, the saturation concentration \(c^L_A = n_A/V_A\) is related to \([S_{A,T}]\) as follows:

\[ [S_{A,T}] = \left( \frac{n_A}{V} \right) \left( \frac{V_A}{V_A} \right) = c^L_A \left( \frac{V_A}{V} \right) = c^L_A \phi_A, \]  

(4.19)

and

\[ [S_{B,T}] = [S_T] - [S_{A,T}] = [S_T] - c^L_A \phi_A. \]  

(4.20)

Here, \(\phi_A\) is the volume fraction of phase \(A\). With these, we can get Equation (4.16) in terms of \(c^L_A\):

\[ [S_{A,T}] = [S_A] + [S_A L] = c^L_A \phi_A; \quad [S_{B,T}] = [S_B] + [S_B L] = [S_T] - c^L_A \phi_A, \]  

(4.21)
By solving this system of equations, we can determine the ligand PC given by:

\[
PC = \left( \frac{\phi_A}{1 - \phi_A} \right) \left( \frac{[S_BL] + (1 - \phi_A)[L]}{[S_AL] + \phi_A[L]} \right).
\]  

(4.22)

Following the same procedure, for a system where ligand binding to the scaffold is described by a second order polynomial in both phases, the system of equations that describe each binding reaction is given by:

\[
k_A = \frac{[S_AL]}{2[S_A][L]}; \quad k_B = \frac{[S.BL]}{2[S_B][L]},
\]  

(4.23)

\[
k_A^2 = \frac{[S_AL]}{[S_A][L]^2}; \quad k_B^2 = \frac{[S.BL]}{[S_B][L]^2},
\]  

(4.24)

\[
[S_{A,T}] = [S_A] + [S_AL] + [S_AL^2] = c_A^L \phi_A,
\]  

(4.25)

\[
[S_{B,T}] = [S_B] + [S.BL] + [S_BL^2] = [S_T] - c_A^L \phi_A,
\]  

(4.26)

and

\[
[L_T] = [L] + [S_AL] + 2[S_AL^2] + [S_BL] + 2[S_BL^2].
\]  

(4.27)

Here, \([S_AL^2]\) and \([S_BL^2]\) are the concentrations of scaffolds with two ligands bound in phase A and B, respectively. Furthermore, when ligand binding can be described by a second order polynomial,
\[ c_A = c_A \left( \frac{P_A}{P_B} \right) = c_A \left( \frac{1 + 2k_A[L] + k_A^2[L]^2}{1 + 2k_B[L] + k_B^2[L]^2} \right), \tag{4.28} \]

and finally we get

\[ PC = \left( \frac{\phi_A}{1 - \phi_A} \right) \left( \frac{[S_B L] + 2[S_B L_2] + (1 - \phi_A)[L]}{[S_A L] + 2[S_A L_2] + \phi_A} \right). \tag{4.29} \]

### 4.4 Results

#### 4.4.1 Overall Effects Of Different Types Of Ligands

The simulations allow us to probe the effects of ligand binding on the low- and high-concentration arms of scaffolds. Figure 4.3C shows binodals for the scaffold without ligand and the impact of ligand binding on the binodals. In each case, we show results for a ratio of 0.23 for the ligand to scaffold sites. For this ratio, we observe four categories of ligand-modulated phase behaviors of scaffolds: 1) Ligand binding abolishes scaffold phase separation (phase diagrams with red bounding boxes in Figure 4.3C); 2) ligand binding destabilizes scaffold phase separation (phase diagrams with blue bounding boxes in Figure 4.3C); 3) ligand binding does not impact phase separation (phase diagrams with purple bounding boxes in Figure 4.3C); and 4) ligand binding promotes phase separation (phase diagrams with green bounding boxes in Figure 4.3C). Results for three additional ligand-to-scaffold ratios are shown in Figure 4.4, Figure 4.5, and Figure 4.6, and the results are qualitatively similar to those shown in Figure 4.3C.
Figure 4.3: Measuring how ligands affect scaffold phase behavior (ligand to scaffold ratio of 0.23). (A) Schematic of the scaffold molecule used in coarse-grained simulations. The molecule has five sticker sites and two explicit spacer sites (the remaining two spacer sites are implicit). (B) Three energy scales were examined to assess how the different ligands modulate scaffold phase separation. (C) Binodals of the scaffold in the absence of ligand (gray) and in the presence of the given ligand (orange) for the three different energy scales, $E_1$, $E_2$, and $E_3$. The bounding boxes for each case are color coded to summarize the effect of each ligand on scaffold phase separation: ligand binding abolishes phase separation (red), ligand binding destabilizes phase separation (blue), ligand binding does not change phase separation (purple), and ligand binding promotes phase separation (green).
Monovalent ligands destabilize or abolish scaffold phase separation, and this is true irrespective of whether they interact with sticker or spacer sites (columns 1 and 2 of Figure 4.3C). Divalent ligands that interact directly with sticker sites on scaffolds also destabilize phase separation by competing with the sticker–sticker interactions that drive phase separation (column 3 of Figure 4.3C). In contrast, phase separation is stabilized by divalent ligands that interact with scaffold spacer sites (column 4 of Figure 4.3C). Bipartite divalent ligands that bind both sticker and spacer sites of scaffolds show an intermediate effect compared to the other two divalent ligands. They can promote phase separation at higher interaction strengths; however, this effect is weaker when compared to that of divalent ligands that bind only to spacer sites (column 5 of Figure 4.3C).
Figure 4.4: Measuring how ligands affect scaffold phase behavior (ligand to scaffold ratio of 0.11). Effect of ligand types on scaffold phase behavior for a ratio of 0.11 ligand to scaffold sites. Binodals of the scaffold in the absence of ligand (grey) and in the presence of the given ligand (orange) for the three different energy scales, $E_1$, $E_2$, and $E_3$. The bounding boxes for each case are color coded to summarize the effect of each ligand on scaffold phase separation: red – ligand binding abolishes phase separation; blue – ligand binding destabilizes phase separation; purple – ligand binding does not change phase separation; and green – ligand binding promotes phase separation.
Figure 4.5: Measuring how ligands affect scaffold phase behavior (ligand to scaffold ratio of 0.34). Effect of ligand types on scaffold phase behavior for a ratio of 0.11 ligand to scaffold sites. Binodals of the scaffold in the absence of ligand (grey) and in the presence of the given ligand (orange) for the three different energy scales, $E_1$, $E_2$, and $E_3$. The bounding boxes for each case are color coded to summarize the effect of each ligand on scaffold phase separation: red – ligand binding abolishes phase separation; blue – ligand binding destabilizes phase separation; purple – ligand binding does not change phase separation; and green – ligand binding promotes phase separation.
Figure 4.6: Measuring how ligands affect scaffold phase behavior (ligand to scaffold ratio of 0.46). Effect of ligand types on scaffold phase behavior for a ratio of 0.11 ligand to scaffold sites. Binodals of the scaffold in the absence of ligand (grey) and in the presence of the given ligand (orange) for the three different energy scales, $E_1$, $E_2$, and $E_3$. The bounding boxes for each case are color coded to summarize the effect of each ligand on scaffold phase separation: red – ligand binding abolishes phase separation; blue – ligand binding destabilizes phase separation; purple – ligand binding does not change phase separation; and green – ligand binding promotes phase separation.

4.4.2 Effects Of Ligands On Dilute Phase Concentrations

Experimental characterizations of full binodals are challenging, and accordingly, these measurements have been performed only for a small number of scaffold molecules [37–40]. In contrast, it is easier to measure changes in saturation concentrations in the absence [6] and the
presence of ligands [15, 41]. To set up expectations regarding how saturation concentrations change, we analyze changes as a function of ligand concentration for each of the five ligand types (Figure 4.7). At a given ligand concentration, the greater the difference between $c_L^A$ and $c_A$, the greater the asymmetry in the preferential binding of the ligand to the scaffold in either the dense or dilute phase.

For all ligand concentrations and interaction strengths tested here, both monovalent ligands and the divalent ligand that interacts with scaffold stickers cause an increase in $c_L^A$ thereby destabilizing scaffold phase separation. The extent of destabilization increases monotonically as the energy scale is increased for the monovalent ligands (see magenta and orange traces in panels A-C of Figure 4.7). However, the destabilizing effects of the divalent ligand that interacts with scaffold stickers changes non-monotonically with the energy scale. For a fixed ligand concentration, the extent of destabilization increases upon doubling the ligand-scaffold interaction energy from $E_1$ to $E_2$. However, the extent of destabilization then decreases upon further doubling of the ligand-scaffold interaction energy from $E_2$ to $E_3$. Multivalent ligands that bind to scaffold stickers tend to destabilize phase separation by competing with sticker-sticker interactions. However, at higher ligand-scaffold interaction strengths, the extent of destabilization can be reduced because the system now uses ligand-mediated crosslinks. Therefore, the interplay between ligand valence and the relative strengths of scaffold-scaffold versus ligand-scaffold interactions can be modulated to obtain non-monotonic changes to condensate stability.
The divalent ligand that interacts with scaffold spacers promotes scaffold phase separation as evidenced by the fact that decreases $c_L^A$ vis-à-vis $c_A$ for all ligand concentrations and interaction strengths tested here. We observe a weak non-monotonic trend in that $c_L^A$ decreases and then increases with increasing ligand concentration. A similar effect has been reported in experiments that characterized the phase behavior of a poly-SH3-poly-PRM system in the presence of increased concentrations of the ligand heparin [28]. For this system, the non-monotonic behavior was hypothesized to be due to electrostatic repulsions of heparin at high concentrations. Since charge effects are not included in our model, the simplest explanation is that the non-monotonic behavior results from a ligand concentration dependent interplay between scaffold-scaffold interactions being the only drivers of phase separation to some of these scaffold-scaffold interactions being competed out by ligand-scaffold interactions. The latter is a consequence of increased ligand concentration, which means that it becomes more likely to make ligand-scaffold interactions when compared to scaffold-scaffold interactions.
For the bipartite divalent ligand that interacts with sticker and spacer sites on scaffolds the ligand causes minimal changes to when compared to $c_A$ for the lowest interaction strength, $E_1$. However, there is a modest destabilization of scaffold phase separation at the highest ligand concentrations (Figure 4.7A). At higher interaction strengths and low ligand concentrations, the ligand promotes phase separation, but to a weaker extent than the divalent ligand that interacts purely with spacer sites on scaffolds. Again, when the bipartite divalent ligand promotes scaffold phase behavior, we observe non-monotonic behavior in the dependence of on ligand concentration. This non-monotonic behavior is a general feature of ligands that promote phase separation by preferentially binding to scaffolds in the dense phase.

The effects of bipartite ligands can be further tuned by changing the relative strength of the interaction between the site that interacts with scaffold stickers and the site that interacts with scaffold spacers (Figure 4.8). When the interaction strength with the scaffold spacer site is stronger than the interaction strength with the scaffold sticker site scaffold phase separation is promoted. In contrast, when the interaction strength with the scaffold sticker site is stronger than the interaction strength with the scaffold spacer site scaffold phase separation is destabilized. These results show that modulating the relative sticker versus spacer interaction strengths within a ligand provides an additional handle for modulating ligand mediated control of scaffold phase behavior.
4.4.3 Effects Of Ligands On Dense Phase Concentrations

We assessed how the scaffold and total dense phase concentrations in the presence of each of the ligand types change compared to the behavior of the scaffold alone (Figure 4.9). The key observations are as follows: The scaffold concentration in the dense phase does not generally increase above the ligand-free case regardless of how ligand binding influences $c_A$ (Figure 4.9A-C). Ligands, irrespective of whether they enhance or weaken phase separation, will have a diluting effect on scaffolds within the dense phase and this effect increases with...
increasing ligand concentration (Figure 4.9A-C). Dilution of the dense phase upon increasing ligand concentration has been observed experimentally for hnRNPA1 with BSA as a ligand [42]. The extent of dilution depends on the interaction mode, interaction strengths, and ligand concentration. Specifically, binding to sticker sites on scaffolds has a greater effect on reducing the scaffold concentration in the dense phase, and this effect is increased as the interaction strength between the scaffold and ligand is increased. Further, ligands that do not change $c_A$ can still modulate dense phase properties by reducing the scaffold concentration in the dense phase. The structural consequences, whereby ligands dilute the concentrations of scaffolds within the dense phase, follow from the requirement that the scaffolds have to accommodate ligands within condensates.
Figure 4.9: Changes in dense phase versus dilute phase concentrations as a function of ligand type, interaction strength, and ligand concentration. Top row shows the change in the scaffold concentration in the dense phase for the energy scales (A) $E_1$, (B) $E_2$, and (C) $E_3$. Bottom row shows the change in the total concentration of scaffolds and ligands in the dense phase for the energy scales (D) $E_1$, (E) $E_2$, and (F) $E_3$. The size of each point increases with ligand concentration. Data are plotted only if the system undergoes phase separation, i.e., the width of the two-phase regime satisfies the criterion $(c_B^L - c_A^L) > 0.15$.

Next, we assessed the extent to which dilution of the scaffold concentration within condensates is compensated by an increase in ligand concentration. We quantified the total concentration of scaffolds and ligands in the dense phase for each of the different ligand types. The results, shown in Figure 4.9D-F, may be summarized as follows: ligands that promote phase separation tend to increase the total dense phase concentration (blue); ligands that do not alter the driving forces for phase separation tend to maintain the total dense phase concentration (green); and ligands that destabilize phase separation tend to decrease the total dense phase concentration (red).
4.4.4 Effects Of Ligands On Condensate Structures

To uncover a “molecular” level understanding of the observations summarized above, we quantified site-to-site radial molecularity profiles, which we denote as \( N(r_i) \) and define as the number of sticker or ligand occupied lattice sites in a shell that lies in the interval \((r_i, r_i + \Delta)\) from each scaffold sticker site in the simulation volume. Details of how \( N(r_i) \) profiles are calculated using pair distribution functions are provided in Sub-section 4.3.4.

Row A in Figure 4.10 shows scaffold sticker-to-sticker radial molecularity profiles plotted for \( r \leq L/2 \), where \( L \) is the length of the side of the cubic simulation box. The short-range peak is a signature of phase separation [23]. Destabilization of condensates leads to a ligand concentration dependent decrease and eventual abrogation of the first peak. This is seen in the left three panels of row A of Figure 4.10. The monovalent ligands that bind only to spacer sites of scaffolds cause a dilution of sticker-to-sticker contacts. This is realized by enhancing the effective excluded volume of spacers and weakening the cooperativity of the inter-sticker crosslinks needed for driving phase separation (compare Figure 4.10C, panels 1 and 3). Row B of Figure 4.10 shows the sticker-to-ligand radial molecularity profiles. For destabilizing ligands, a short-range peak is either present only at low ligand concentrations or non-existent as seen in the three left most panels of row B of Figure 4.10. This is the result of ligand-mediated destabilization of condensates and the formation of sticker-ligand interactions in the dispersed one-phase regime.
Figure 4.10: Analysis of site-site radial molecularity profiles for ligand-scaffold energy scale $E_2$.

(A) Sticker-to-sticker radial molecularity profiles quantifying how scaffold stickers organize around one another for different ligand concentrations and types. The dashed profiles are for ligand-free cases. The presence of a peak in the profile is indicative of phase separation, and the heights of these peaks, which change with ligand concentration, quantify the effects of ligands on condensate stability. Monovalent ligands and divalent ligands that interact with stickers, weaken the sticker–sticker interactions, diluting stickers around one another and thereby destabilize and/or dissolve condensates. (B) Sticker-to-ligand radial molecularity profiles that quantify the organization of ligands around scaffold sticker sites for different ligand types and ligand concentrations. The preferential accumulation of ligands around sticker sites in dense phases is evident in the growth of the height of the first peak with increasing ligand concentration. (C) A schematic summarizing the effects of ligands on the structural organization of scaffold stickers. The first box is in the absence of ligands. The insets for each box depict the various interactions occurring for that system. The blue sites denote scaffold stickers and red sites denote scaffold spacers. Scaffold sticker-sticker interactions are destabilized for the monovalent ligands and for the divalent ligand that binds scaffold stickers. The divalent ligand that binds scaffold spacers and the bipartite divalent ligand maintain scaffold sticker–sticker interactions and also provide additional networking interactions. These interactions for the bipartite ligand cause an increase in the correlation length between scaffold stickers within the dense phase.
For divalent ligands that bind directly to spacers and bipartite ligands that bind to spacers and stickers, we observe a maintenance of the first peak in both sets of radial molecularity profiles at all ligand concentrations (see right two panels in rows A and B of Figure 4.10). The height of the first peak of the sticker-to-sticker radial molecularity profile shows non-monotonic behavior. The non-monotonic behavior can be explained by the increase in the height of the first peak in the sticker-ligand radial molecularity profile as ligand concentration increases (see right two panels in row B of Figure 4.10). Accordingly, as ligand concentrations increase, ligand-mediated crosslinking become auxiliary drivers of scaffold phase separation.

The bipartite divalent ligands cause a dilution of scaffolds within the dense phase as shown in Figure 4.9B. This can be explained by the radial molecularity profiles. There is a slight increase in the first peak of the sticker-to-sticker radial molecularity profile vis-à-vis the ligand-free case (last panel in Figure 4.10A). This increase is followed by a decrease, rightward shift and widening of the first peak of the sticker-sticker profiles. Concurrently, the heights of the first peaks of ligand-to-sticker radial molecularity profiles increase monotonically and also exhibit rightward shifts. The inference is that scaffold sticker-sticker interactions are replaced by scaffold-ligand interactions thereby increasing the correlation length between scaffold stickers (last panel in Figure 4.10C). This implies that at the highest ligand concentrations even though does not change relative to $c_A$, the structural organization of scaffold stickers in the dense phase still changes due to interactions with the ligand. These alterations to the structural organization of scaffolds within condensates points to an additional regulatory function that ligands can exert over condensates. The impacts of ligands on structural organization of scaffold sites within condensates should be testable via suitable scattering experiments in vitro [34] or super resolution [43] based measurements in cells.
4.4.5 *PCs* Of Ligands Alone Cannot Predict Which Phase Ligands Preferentially Bind To

To determine if *PCs* alone could be used to determine whether a ligand binds preferentially to the dilute or dense phase, we calculated the *PCs* for the divalent ligands at $T^* \approx 1.08$. The ligand *PC* is defined as the concentration of the ligand in the dense phase divided by the concentration of the ligand in the dilute phase. Panels A and B in Figure 4.11 show that there is a general trend for *PCs* to increase as ligands increasingly promote phase separation. However, when we examine a particular ligand concentration, we find that the *PC* for ligands that destabilize phase separation can be greater than *PCs* that correspond to ligands that promote phase separation (Figure 4.11C). This behavior depends, at least partially, on the strength of the interaction between the ligand and the scaffold. These results suggest that the rank ordering of the *PCs* is not useful for discerning the effects of ligand on scaffold phase behavior.
Figure 4.11: Comparing ligand PC's to the modulation of scaffold saturation concentration. (A) The changes in saturation concentration for all divalent ligands at $T^* \sim 1.08$ are plotted against the corresponding PC's. The size of each circle is proportional to the ligand concentration. The gray box denotes a region of PC's that is consistent with ligands that destabilize, do not change, or promote scaffold phase separation. Data are plotted only if the system undergoes phase separation, i.e., the width of the two-phase regime satisfies the criterion $(c_B^L - c_A^L) > 0.15$. (B) Boxplots of the data presented in (A). (C) PC's as a function of ligand concentration for each of the divalent ligands and energy scales examined in this paper. (D) PC's measured by Ghosh et al. [28](26), colored by the effect of the ligand on scaffold phase behavior at the same concentration. In all plots, blue implies that the ligand promotes scaffold phase separation, green implies that the ligand does not change the drive to phase separate, and red implies that the ligand destabilizes scaffold phase separation.
Figure 4.11A-B also show that there is a large range of ligand PC values that can correspond to all three modulatory effects of ligands on phase behavior. Specifically, we find that PC values spanning from 10 to 100 can correspond to ligands that destabilize, do not change, or promote scaffold phase separation. Additionally, all PCs are found to be greater than one. We also find that for a given ligand, PCs decrease with increasing ligand concentration regardless of the ligand modulatory effect on scaffold phase behavior (Figure 4.11C). Together these results suggest that PCs are not a direct measure of preferential ligand binding. This is because PCs of ligands, unlike those of scaffolds, are a convolution of factors including ligand concentrations and interaction strengths of ligands for scaffold sites.

Ghosh et al., [28] recently reported a complete assessment of ligand PCs and their effects on scaffold phase behavior. Focusing on the poly-SH3:poly-PRM system, Ghosh et al., measured how lysozyme and heparin modulated the saturation concentration of poly-SH3:poly-PRM condensates. At high concentrations, lysozyme destabilizes the formation of poly-SH3:poly-PRM condensates. In contrast, heparin promoted the formation of poly-SH3:poly-PRM condensates at low and intermediate heparin concentrations, but destabilized condensate formation at high heparin concentrations. Additionally, Ghosh et al., measured PCs at three different ligand concentrations where they quantified the effects of ligands on condensate stability. Figure 4.11D summarizes the key takeaways from these experiments. The bar color indicates the effect on condensate formation, where blue indicates promotion of phase separation, green indicates no change in phase behavior, and red indicates destabilization of phase separation. Both lysozyme and heparin show a decrease in PC as the ligand concentration is increased. Our simulation results are in agreement with these observations from experiments. Further, the data of Ghosh et al., show that PCs of ligands that promote phase separation can be lower than those of ligands that destabilize phase separation, even though the experiments were performed at similar bulk concentrations of ligands.
4.4.6 Polyphasic Linkage Theory Establishes That PCs Of Ligands Combine Contributions Of Preferential Binding And Local Concentration Effects

The impact of ligand binding on saturation concentrations is written in terms of binding polynomials $P_A$ and $P_B$ that quantify the binding of the ligand in question to the scaffold in phases $A$ and $B$, respectively. A binding polynomial is the partition function of the ligand plus scaffold system and is a sum over the activities of all states in the system involving the scaffold relative to the free scaffold [36](36). We assume two types of systems, one where ligand binding to the scaffold is described by a first order polynomial in both phases and one where ligand binding to the scaffold is described by a second order polynomial in both phases. When binding can be described by a first order polynomial,

$$c_A^L = c_A P_A = c_A \frac{1 + k_A[L]}{1 + k_B[L]},$$  \hspace{1cm} (4.30)

where $[L]$ is the free ligand concentration and $k_A$ and $k_B$ are the association constants of the ligand to the scaffold in phase $A$ and $B$, respectively. Likewise, when binding can be described by a second order polynomial,

$$c_A^L = c_A \frac{P_A}{P_B} = c_A \frac{1 + 2k_A[L] + k_A^2[L]^2}{1 + 2k_B[L] + k_B^2[L]^2},$$  \hspace{1cm} (4.31)

In its simplest form, linkage theory assumes that the dense phase scaffold concentration does not change in the presence of ligand. From our coarse-grain simulations, we find that this assumption is only reasonable for systems with ligands that promote phase separation with a ligand to scaffold molecule ratio of less than two (Figure 4.9). Therefore, we focused our
analysis on systems with $k_B/k_A = 2, 4, 6, 8, \text{ and } 10$ and total ligand to scaffold concentration ratios spanning 0.25 to 1.75. The first criterion imposes preferential dense phase binding and hence the promotion of phase separation.

We examined the relationship between $c^L_A$ and $PC$ by solving for the system of equations that describe each binding reaction and using the fact that the total scaffold concentration in phase $A$ is given by $c^L_A \phi_A$, where $\phi_A$ is the volume fraction of phase $A$ (see SI Appendix for details). In both phases the ligand can be free or bound, and the partitioning of free ligand between the two phases is governed by the relative volumes of each phase. Therefore, when binding is described by first order polynomials in both phases, it follows that

$$PC = \left( \frac{\phi_A}{1 - \phi_A} \right) \left( \frac{[S_B L] + (1 - \phi_A)[L]}{[S_A L + \phi_A[L]]} \right), \quad (4.32)$$

where $[S_A L]$ and $[S_B L]$ are the concentrations of the bound scaffold in phase $A$ and $B$, respectively. Likewise, when binding is described by second order polynomials in both phases,

$$PC = \left( \frac{\phi_A}{1 - \phi_A} \right) \left( \frac{[S_B L] + 2[S_B L_2] + (1 - \phi_A)[L]}{[S_A L_2] + 2[S_A L_2] + \phi_A[L]} \right), \quad (4.33)$$

where $[S_A L_2]$ and $[S_B L_2]$ are the concentrations of the scaffold bound by two ligands in phase $A$ and $B$, respectively.
Figure 4.12: Theory shows how $PC$s and scaffold saturation concentration changes as a function of ligand concentration for ligands with different binding modes. Ligands that bind as described by a first-order polynomial are shown in the blue-to-green color scale, and ligands that bind as described by a second-order polynomial are shown in the orange-to-yellow color scale. (A) The change in scaffold saturation concentration as a function of the ratio of the total ligand concentration to the total scaffold concentration. (B) The $PC$ as a function of the ratio of the total ligand concentration to the total scaffold concentration. (C) The $PC$ as a function of the change in the scaffold saturation concentration for a ligand to scaffold ratio of 0.8.

Figure 4.12 shows how $PC$s change as a function of total ligand to scaffold ratio, $k_B/k_A$, and whether binding is first or second order. We set $c_A = 1 \mu M$, $c_B = 19 \mu M$, and the total scaffold concentration, $[S_T] = 10 \mu M$. As designed, all systems show a decrease in compared to $c_A$, implying phase separation is promoted upon ligand binding (Figure 4.12A). Consistent with our coarse-grained simulation results and the results of Ghosh et al., we observe that $PC$s decrease and approach one at high ligand concentrations for all systems (Figure 4.12B). However, the slopes and details of how and $PC$ change with ligand concentration vary depending on the binding mode.
4.4.7 Effects Of Ligands On Two-Component Systems With Heterotypic Interactions Only

Up to this point, we have primarily focused on systems with only one scaffold species, such that the driving forces for phase separation become effectively homotypic; albeit without any intra-molecular interactions. In systems where heterotypic interactions are the main drivers of phase separation, the polyphasic linkage formalism requires generalization, since the contributions to binding polynomials involve multiple species. This can be illustrated using new sets of preliminary simulations where we model phase separation as being the result of purely heterotypic interactions and assess the impact of ligands on this type of phase behavior. Here, we now have multiple components where, under conditions that generate two phases, each component has it’s corresponding coexisting concentrations. Therefore, we shall adopt a modified notation for this Sub-section, and the next.

Consider two scaffold macromolecules A and B. If we have two co-existing phases, then the dilute phase is \( c_{\text{dil}} \), the dense phase is \( c_{\text{den}} \). Furthermore, in the presence of the ligand, each phase can be modulated and is represented as \( c_{\text{den},L} \). Lastly, a superscript for the coexisting concentrations shall now refer to the concentration of that particular scaffold: e.g., \( c_{\text{den},L}^A \) is the concentration of scaffold A in the dense phase, where a ligand is also present in the system.

Figure 4.13 shows how the inclusion of a ligand modulates purely heterotypic phase separation as assessed in a lattice simulation [23]. Each scaffold has seven sticker sites that makes a single reversible cross-link with a site on the other scaffold (Figure 4.13A). The phase boundary for this system shows closed loop behavior as expected for purely heterotypic interactions (Figure 4.13B, black). Here, [scaffold A] and [scaffold B] refer to the total concentrations, in units of stickers per lattice sites, of scaffold molecules A and B, respectively. To assess
the effect of a ligand on overall phase behavior, we included a ligand with two binding sites that are identical to the stickers of scaffold A (Figure 4.13A). Such a system mimics the scaffold-client systems studied by Banani et al., [11], although the concentrations of client molecules in their work were low enough to ensure that the client does not behave like a modulatory ligand.
Figure 4.13: Example effect of a ligand on the phase behavior of linear multivalent proteins with purely heterotypic interactions. (A) Schematic of model system. Each scaffold has seven sticker sites. The ligand has two sticker sites, which are the same as scaffold A and thus can only bind scaffold B. (B) Phase boundaries of the system without (black, stars) and with (purple, circles) the ligand shown in log and linear scale. Here, concentration is in stickers/lattice sites. Log scale allows for viewing changes in $c_{\text{dil}}$, whereas linear scale allows for viewing changes in $c_{\text{den}}$. The color of the circles for the ligand case correspond to the ratio of the input concentrations of the scaffolds, $[\text{scaffold A}]/[\text{scaffold B}]$. When the input concentration has an excess of scaffold A, the color is pinker, and when the input concentration has an excess of scaffold B, the color is more blue-green. All sticker-sticker interaction energies were set to $2/T^*$, where $T^*$ is the effective temperature. The ligand concentration was set to $1 \times 10^{-3}$ molecules/lattice sites. (C) Change in $c_{\text{dil}}$ and $c_{\text{den}}$ for each scaffold molecule as a function of the ratio of the input concentrations, $[\text{scaffold A}]/[\text{scaffold B}]$. If $c_{\text{dil},L}/c_{\text{dil}}$ is less than one, then the ligand promotes phase separation of that scaffold, whereas if $c_{\text{dil},L}/c_{\text{dil}}$ is greater than one, then the ligand suppresses phase separation of that scaffold. When taking account the shifts for both scaffold A and B, we find that when $[\text{scaffold A}]/[\text{scaffold B}]$ is large, the ligand slightly suppresses phase separation (light red shade). This suppression increases as $[\text{scaffold A}]$ approaches $[\text{scaffold B}]$ (darker red shade). When $[\text{scaffold B}]$ is slightly larger than $[\text{scaffold A}]$, the ligand has the greatest effect at promoting phase separation (dark blue shade). When $[\text{scaffold B}]$ is much larger than $[\text{scaffold A}]$, then the ligand slightly promotes phase separation (light blue shade). In regard to the dense phase, the concentration of scaffold B in the dense phase shows limited change in the presence of the ligand, whereas the concentration of scaffold A in the dense phase decreases as $[\text{scaffold A}]/[\text{scaffold B}]$ decreases. (D) Schematic summarizing the results shown in panels (A)–(C).

Our simulations show that the addition of a low-valence ligand can suppress or promote scaffold phase separation depending on the relative total concentrations of scaffold A and scaffold B (Figure 4.13B–C). In the presence of a large excess of scaffold A, the ligand can suppress phase separation as shown by the observation that $\left(\frac{c_{\text{dil},L}}{c_{\text{dil}}}\right) > 1$ (Figure 4.13C, light red). This is because the sub-stoichiometric amount of scaffold B is partially sequestered by interactions with the ligand, thus requiring a higher overall concentration of scaffold B to drive phase separation (Figure 4.13C, squares, light red).
For a two-scaffold system with two scaffolds of equal effective valence that undergoes phase separation purely via obligate heterotypic interactions, such as the system of study here, $c_{\text{dil,L}}^A$ and $c_{\text{dil,L}}^B$ are minimized for a 1:1 stoichiometry of the scaffolds. This arises when $[\text{scaffold A}] = [\text{scaffold B}]$ (Figure 4.13B, black).27 Ligands with two binding sites that are identical to the stickers of scaffold A have the largest effect on modulating phase separation when $[\text{scaffold A}] = [\text{scaffold B}]$ (Figure 4.13C, dark red and dark blue). Accordingly, the relevant parameter is the input stoichiometric ratio of scaffolds A and B, which we define as $r = \frac{[\text{scaffold A}]}{[\text{scaffold B}]}$. As $r$ decreases from $\sim 2$ to $\sim 0.8$, the ability of the ligand to suppress phase separation is maximized as evidenced by the fact that $(c_{\text{dil,L}}/c_{\text{dil}}) > 1$ for both scaffolds (Figure 4.13(c), dark red). However, as $r$ decreases below 0.8, the presence of the ligand leads to a promotion of scaffold phase separation whereby $(c_{\text{dil,L}}/c_{\text{dil}}) < 1$ for both scaffolds A and B (Figure 4.13C dark blue). The promotion of phase separation in this regime arises from the fact that the ligand renormalizes the concentration of scaffold A. When there is a large excess of scaffold B, the ligand has a significant modulatory effect on the concentrations of scaffold molecules in the dense phase (Figure 4.13C) Although the presence of the ligand always decreases $c_{\text{den}}^A$ and $c_{\text{dil}}^B$, $c_{\text{den}}^A$ shows the largest decrease when $r < 1$.

When phase separation is driven by heterotypic rather than homotypic interactions, additional variables influence how a ligand regulates scaffold phase separation. These variables include the relative total concentrations of the scaffold molecules and which scaffold molecule the ligand binds. Low-valence ligands are not general suppressors of phase separation as was observed for the case of purely homotypic interactions. For instance, when the ligand is a lower-valence version of one of the scaffolds, the ligand can also engage in heterotypic interactions that can suppress or support networking. Therefore, the ability to suppress or promote phase separation will depend on the relative input concentrations of the scaffold molecules.
4.4.8 Effects Of Ligands On Two-Component Systems With Homotypic And Heterotypic Interactions

We next assessed how ligands impact phase behavior when there is an interplay between homotypic and heterotypic interactions. These results are summarized in Figure 4.14. Here, scaffold A can also engage in homotypic interactions (Figure 4.14A) and the interaction strengths of the homotypic and heterotypic interactions are set to be equal. In the absence of a ligand, the phase boundary is no longer a closed loop (Figure 4.14B) black stars]. This is because when scaffold A is in excess, scaffold A can still phase separate through homotypic interactions. We assess the effects of two types of ligands on the joint phase behavior of the two scaffolds. The ligands considered include a divalent ligand with sites that are identical to those of the stickers on scaffold A. This is denoted as ligand A. The ligand designated as B is also divalent, and the binding sites are identical to those of the stickers on scaffold B (Figure 4.14A). When \( r > 1 \) and the total concentration of scaffold A is greater than that of scaffold B, ligands A and B have similar effects on the phase behavior of the scaffolds (Figure 4.14B-C). Specifically, \( c_{A}^{dil} \) changes minimally whereas \( c_{B}^{dil} \) increases (Figure 4.14C). The implication is that neither of the ligands have a significant influence on the homotypic interactions among scaffold A molecules. In contrast, both ligands reduce the ability of scaffold B to co-phase separate. This is because ligand A interacts preferentially with scaffold B in the dilute phase and sequesters it from the dense phase, whereas ligand B competes with scaffold B for heterotypic interactions with scaffold A in the dense phase, leading to an upshift in \( c_{dil}^{B} \).
Figure 4.14: Example effect of a ligand on the phase behavior of linear multivalent proteins with homotypic and heterotypic interactions. (A) Schematic of model system. Each scaffold has seven sticker sites. Scaffold A can interact with itself or with scaffold B, whereas scaffold B can only engage in heterotypic interactions with scaffold A. The two ligands examined were divalent versions of scaffold A and scaffold B. (B) Phase boundaries of the system without (black, stars) and with ligand A (filled-in circles) and with ligand B (empty circles) shown in log and linear scale. Here, concentration is in units of stickers/lattice sites. The log scale allows for viewing changes in $c_{\text{dil}}$, whereas the linear scale allows for viewing changes in $c_{\text{den}}$. The color of the circles for simulations in the presence of ligand correspond to the input stoichiometric ratio $r$. When the input concentration involves an excess of scaffold A, the color is pinker, and when the input concentration involves an excess of scaffold B, the color is blue-green. All sticker-sticker interaction energies were set to $2/T^*$, where $T^*$ is the effective temperature. The ligand concentration was set to $1 \times 10^{-3}$ molecules/lattice sites. (C) Changes in $c_{\text{dil}}$ and $c_{\text{den}}$ for each scaffold molecule as a function of the ratio of $r$. If $c_{\text{dil},L}/c_{\text{dil}}$ is less than one, then the ligand promotes phase separation of that scaffold, whereas if $c_{\text{dil},L}/c_{\text{dil}}$ is greater than one, then the ligand suppresses phase separation of that scaffold. (D) Summary schematic of the above results.

The effects of the two ligands diverge when $r$ is less than one (Figure 4.14B-C). When the total concentration of scaffold B is greater than that of scaffold A, ligand A promotes scaffold phase separation, whereas ligand B suppresses phase separation (Figure 4.14C). The presence of ligand A reduces $c_{\text{dil}}^A$ and $c_{\text{dil}}^B$, thereby promoting phase separation of both scaffolds. In contrast, the presence of ligand B leads to an increase in $c_{\text{dil}}^A$, implying that ligand B sequesters scaffold A in the dilute phase. As the total concentration of scaffold B increases with respect to that of scaffold A, the degrees to which ligands A and B respectively promote vs suppress phase separation decrease (Figure 4.14C). Additionally, we find that both ligand types reduce the concentration of scaffolds in the dense phase (Figure 4.14B-C). Ligand A always decreases $c_{\text{den}}^B$ to a greater extent than ligand B. The largest decrease in scaffold concentration in the dense phase is observed for scaffold A in the presence of ligand A; see results for $c_{\text{den}}^A$ in the presence of ligand A, when $r < 1$, i.e., $[\text{scaffold A}] < [\text{scaffold B}]$. Given the excess of scaffold
B, phase separation is dominated by heterotypic interactions, and this decrease in $c_{\text{den}}^A$ mimics the observations for the purely heterotypic case shown in Figure 4.13.

4.5 Discussion

The stickers-and-spacers formalism [29, 44–47] allows us to uncover key features of ligands that destabilize or stabilize phase separation via preferential binding to scaffolds in the dilute versus dense phase, respectively. Overall, our findings are as follows: Monovalent ligands weaken phase separation either by reducing the overall valence of the scaffold when they interact directly with sticker sites or by enhancing the effective excluded volume of spacers and weakening the cooperativity of the inter-sticker crosslinks that is needed for driving phase separation when they interact directly with spacer sites. Divalent ligands that bind to scaffold sites weaken phase separation by competing directly with inter-sticker interactions. In contrast, divalent ligands that bind to spacer sites enable additional networking of multivalent scaffold molecules by serving as crosslinkers, thereby promoting phase separation. This shows that ligands can lower the saturation concentrations for scaffolds, a finding that is important in light of an ongoing debate about the relevance of phase separation in vivo, especially at endogeneous expression levels [20].

Our findings, and those of others [29], imply that scaffolds can undergo ligand-mediated phase separation even if the endogeneous concentration of the scaffold is below its intrinsic saturation concentration. This feature is likely to be amplified by the collective contributions of networks of ligands, providing they are multivalent [10–12]. Finally, bipartite divalent ligands that bind both stickers and spacers within the scaffold can modulate scaffold phase behavior in either direction depending on the relative interaction strengths of the ligand with the stickers and spacers of the scaffold. Ligand-modulation of condensate stability may be
thought of as being another component of heterotypic buffering, a concept recently introduced to describe how the interplay between homotypic and heterotypic scaffold-scaffold interactions regulates scaffold phase behavior in vivo [18].

We find that the concentrations of scaffolds within the dense phase stays similar to that of the unliganded case or decreases in the presence of ligand. This is true irrespective of whether the ligand binds preferentially to the scaffold in its dense or dilute phase. For preferential binding to the scaffold in the dense phase, $c_A^L$ decreases, and the scaffold concentration in the dense phase ($c_B^L$) generally decreases when compared to $c_B$. This helps accommodate ligands that bind preferentially to scaffold sites in the dense phase. Conversely, for preferential binding to the scaffold in the dilute phase, $c_A^L$ increases and $c_B^L$ decreases compared to $c_B$. This dilution of scaffold within the dense phase derives from weakening the driving forces for phase separation.

We also focused on the important question of how one might quantitatively assess the contributions of ligands as modulators of condensate formation and dissolution in vivo or in vitro. One approach would be to measure $PC$s of ligands since they quantify the enrichment or depletion of ligands in condensates [48]. However, we show that $PC$s of ligands are convoluted quantities that do not provide direct assessments of the effects of ligands as modulators of scaffold phase behavior. Instead, the simplest approach would be direct measurements of scaffold concentrations in the dilute and dense phases as a function of ligand concentration [41].

It is worth noting that we have made the simplifying assumption that a scaffold will be defined by a fixed saturation concentration. However, this is only true if homotypic interactions among scaffold molecules are the primary drivers of phase separation [8, 23]. If condensates form via a combination of homotypic and heterotypic interactions [8, 23, 49], then the
network of these interactions [9, 14] and hence a combination of scaffold concentrations will determine the location of the phase boundary. In this scenario, one would have to measure the effects of ligands on the location of the phase boundary, governed jointly by the concentrations of all scaffold molecules that drive phase separation. This requires measuring the concentrations of more than one scaffold molecule while titrating the concentration of the ligand in question. Further complexities will arise as we consider how a set of distinct ligands impact the phase behavior of condensates that are governed by a network of homotypic and heterotypic interactions of scaffold molecules.

It is also known that the driving forces for phase separation can be modulated by anchoring scaffolds to surfaces [50] or via physical interactions with soft surfaces in cells [51]. For example, Morin et al., [51] showed that preferential interactions of a pioneer transcription factor KLF4 with the surface of DNA can lower the threshold concentration for phase separation. Morin et al., used the Brunauer-Emmett-Teller theory [52] for multilayer adsorption to explain their results. Taken together, it follows that preferential binding of multivalent ligands to scaffolds in dense phases, this work and that of Ghosh et al., [28] and Espinosa et al., [29], and adsorption of scaffolds to surfaces [53, 54] as shown by Morin et al., [51] can lead to similar effects in terms of lowering the saturation concentrations of scaffold macromolecules. Therefore, we propose that the stabilization of condensate formation via surface interactions derives from preferential adsorption of dense phases through spacer-surface interactions and / or enhancing of sticker-sticker crosslinks. If this proposal is valid, then it follows that a unified theory is likely achievable for describing how the stabilities of condensates are impacted by the bulk phase concentrations of preferentially binding ligands and the surface features of soft interfaces.

Our work suggests that the effects of small molecules (ideally, multivalent ligands) on can be used as part of a chemical biology toolkit [55] to infer the features and internal organization
of scaffolds within condensates. For example, if molecules with certain chemical structures destabilize condensates, then one can infer that a complementary interaction motif in the scaffold is primarily accessible in the dilute phase and thus may be involved in driving phase separation. These inferences have the potential to enable the design of small molecules that modulate scaffold phase behavior in prescribed ways.

Lastly, we find that the precise nature of regulation of collective phase separation in systems with two macromolecular scaffolds will depend on whether phase separation is driven purely by heterotypic interactions or if there is an interplay between homotypic and heterotypic interactions. As a rule, ligands that can contribute additional networking interactions will promote phase separation, whereas ligands that disrupt scaffold networking interactions and cannot compensate for the loss of crosslinks among scaffolds will suppress phase separation. For systems in which phase separation depends, at least partially, on heterotypic interactions, the relative concentrations of the scaffolds and the specificity of ligand sites for scaffold stickers will influence how the ligands affect phase separation. The key finding is that the relative total concentrations of scaffolds will determine how different ligands regulate overall phase behavior.


4.6 References


Chapter 5

Structure-Function Relationships In Mitochondrial Transcriptional Condensates

5.1 Preamble

This chapter is based on the following article currently under review at PNAS: Marina Feric, Azadeh Sarfallah, Furqan Dar, Dmitry Temiakov, Rohit V. Pappu (2022). Structure-Function Relationships In Mitochondrial Transcriptional Condensates, [1]. M.F. and T.M. designed the research. M.F. performed all microscopy experiments and analysis of in vitro droplets and in vivo mitochondrial nucleoids. D.T. and A.S. designed the in vitro transcription experiments and A.S. performed them. F.D. and R.V.P. designed the computational model. F.D. developed and deployed the model, performed the simulations, and analyzed the simulation results.
M.F., F.D., R.V.P. and T.M. wrote the manuscript. All authors discussed results and edited the manuscript.

## 5.2 Introduction

Proteins and nucleic acids form diverse biomolecular condensates, that are proposed to arise via macromolecular phase separation [2–5]. Condensates emerge by demixing of protein, RNA and DNA components from their cellular surroundings to form distinct, non-membrane bound cellular structures. The formation of condensates is typically mediated by multivalent homotypic and heterotypic interactions amongst proteins and nucleic acids [2]. Prominent biomolecular condensates include P-granules and stress granules in the cytoplasm [6, 7] as well as the nucleolus and RNA splicing factor speckles in the nucleus [8, 9]. Within mitochondria, the genome-containing mitochondrial (mt)-nucleoid and RNA processing granules are condensates that appear to form via phase separation [10, 11].

Condensates are enriched in functional components, such as transcription factors or RNA processing factors. Increased concentrations of bioactive macromolecules within condensates is thought to be important for enhancing the rates of key biochemical reactions within condensates [12]. However, the connection between condensate structure and function remains unclear [3]. A major hurdle in elucidating structure-function relationships of condensates has been the difficulty to reconstitute functionally active condensates in-vitro with all the biochemically relevant components. Conversely, in-vivo perturbations, such as mutagenesis or pharmacological disruption, make it challenging to isolate the effects of phase behavior from the functional properties of the affected cellular components.

Condensates are thought to contribute to many cellular functions, including genome organization and transcription [13–15]. Major architectural chromatin proteins such as the linker
histone H1 [16] and the heterochromatin protein HP1α form condensates in-vitro and in-vivo [17, 18]. Phase separation, in different manifestations, has been suggested to contribute to higher-order organization of genomes into domains and compartments [19–21]. In particular, various components of the transcription machinery spontaneously concentrate into condensed phases in the mammalian nucleus, including prominently at sites of super-enhancers [22]. This behavior has been attributed to the intrinsically disordered regions (IDRs) found in many transcription factors and chromatin proteins [23]. IDRs are thought to mediate weak, specific, multivalent protein-protein interactions that give rise to dynamic, non-stoichiometric condensed assemblies [24]. Functionally, the condensation of transcription components is associated with bursts of transcription of RNA [25] that are consequently followed by dissolution of the condensate in an effective feedback loop [26].

5.2.1 The mt-Nucleoid As A Model Functional Condensate

The mitochondrial genome (mtDNA) and its own dedicated gene expression machinery are also organized via phase separation [10]. Human mitochondria contains hundreds of copies of their own 16 kb, circular genome [27] that assemble into mitochondrial nucleoids, which are membraneless, nucleoprotein complexes of ∼100 nm in diameter containing mtDNA and associated proteins [28, 29]. In support of phase separation as a driver of mt-nucleoid organization, the major mt-genome architectural protein TFAM phase separates in-vitro and in-vivo, and combined with mtDNA, forms condensates that recapitulate the behavior of mt-nucleoids in cells [10]. The mt-nucleoids serve as sites of transcription of long, polycistronic mtRNA, which becomes further processed in adjacent RNA granules that are associated with the mitochondrial membrane and are also thought to form via phase separation [11].
The relative simplicity, involving only a small number of minimally required components, makes mitochondrial transcription a unique and tractable model system to probe structure-function relationships in a biologically relevant condensate. Mitochondrial transcription can be reconstituted under soluble conditions with only four components: mtDNA, the single-subunit mitochondrial RNA polymerase POLRMT, and two transcription factors, TFAM and TFB2M [30, 31]. Here, we have reconstituted mitochondrial transcription under condensate-forming conditions in-vitro, and we directly probe functional consequences of the condensate morphologies and their internal physicochemical environments. We demonstrate that the mitochondrial transcription machinery forms multi-phasic, dynamically arrested, non-equilibrium condensates with slow diffusivities, contributing to dampened transcriptional kinetics compared to equivalent reactions in bulk solutions. Importantly, we find that the production of nascent RNA during transcription alters the structure of in-vitro the condensate. Our results demonstrate a close interplay between the structure and functional activity of an archetypal biomolecular condensate.

5.3 Materials & Methods

5.3.1 Cell Culture

HeLa cells (ATCC, CCL-2, Lot. #70000153) were grown at 37°C and with 5% CO2 in Dulbecco’s Modified Eagle Medium (DMEM, Thermo Fisher Scientific, #11960-044) supplemented with 10% Fetal Bovine Serum (Thermo Fisher Scientific, #10437), 1% glutamine (Thermo Fisher Scientific, #25030081), and 1% penicillin/streptomycin (Thermo Fisher Scientific, #15140-122).
5.3.2 siRNA

Silencer Select Pre-Designed siRNA constructs were used to knock-down mitochondrial proteins: TFAM–s14002; POLRMT–s1082; TFB2M–s34561; TEFM–s36219; mTERF1–s15508; mTERF2–s37162; mTERF3–s27208; mTERF4–s43559. Cells were seeded on high precision cover slips (Azer Scientific) in 12-well plates (2.5 × 10^4 cells) or in 96-well plates (2.5 × 10^3 cells) on Day 0. On Day 1, cells were transfected with siRNA (15 nM final) using DharmaFECT transfection reagent (Horizon Discovery) and incubated for 72 hours. As a negative control, Silencer Select Negative Control No. 2 siRNA was used (Thermo Fisher Scientific, ASO2GNFA). As a positive control of siRNA transfection, AllStars Hs Cell Death siRNA (Qiagen, 1027299) was used.

5.3.3 Western Blot

Trypsinized cells were collected from 12-well plates, centrifuged, washed with ice-cold PBS, dissolved in 2X Laemmli Sample Buffer (Bio-Rad), and denatured for ~10 min at 95°C. Samples were stored at -20°C. Protein concentration was estimated using a Bradford assay. Equal protein amounts of each sample were loaded onto a 10-well NuPAGE 4-12% Bis-Tris Protein Gel at 150 V for 1 hour followed by wet transfer onto a membrane at 200 mA for 2 hours. Membranes were blocked with 5% BSA and 1X TBST for ~30 minutes at room temperature and incubated with primary antibodies (TFAM (Sigma, HPA063684), beta-actin (Sigma, A2228), MTERF2 (Thermo Fisher Scientific, PA5-109989), and tubulin-alpha (Bio-Rad, VMA00051)) in 5% BSA and 1X TBST at 4°C overnight followed by three wash steps. Membranes were incubated with HRP secondary antibodies for 1 hour at room temperature followed by three wash steps. Proteins were detected using the Amersham ECL Western Blotting Detection Reagent on a Bio-Rad ChemiDoc Imager and quantified with ImageJ.
5.3.4 mtDNA Depletion

Cells were seeded on Day 0 on coverslips on high precision cover slips (Azer Scientific) in 12-well plates (2.5 × 10^4 cells) or in 96-well plates (2.5 × 10^3 cells). On Day 1, media was supplemented with 1 mM sodium pyruvate (Thermo Fisher Scientific) and 50 µg/ml uridine (Thermo Fisher Scientific) [32](43). Cells were transfected with a plasmid containing the DNase Herpes Simplex Virus UL12.5M185 (pMA4008, Addgene plasmid #70109) using the FuGene transfection reagent and incubated for 72 hours [32](43).

5.3.5 Fixation And Labelling

Cells were incubated with 100 nM MitoTracker Red CMXRos (Thermo Fisher Scientific) for ~15 min at 37°C and 5% CO2. To fix cells, 16% paraformaldehyde (PFA) was diluted 1:2 in PBS to make an 8% solution, which was then diluted 1:2 directly into cell culture media (4% PFA final), and cells were fixed for 20 minutes at room temperature. Cells were washed with PBS and stored at 4°C. To permeabilize cells, cells were treated with 0.1% Triton-X for 10 min and washed with PBS. Coverslips were incubated with primary antibodies (anti-DNA clone AC-30-10, EMD Millipore, #CBL186, and anti-TFAM, Atlas Antibodies, HPA063684) without blocking agents for 1 hour (1:500 dilution) at room temperature in a humidity chamber. Coverslips were then washed three times with PBS. Secondary antibodies were applied at 1:1000 for anti-rabbit-488 (Thermo Fisher Scientific, A11008) or 1:50 for anti-mouse-405 (Thermo Fisher Scientific, A31553) for 1 hour at room temperature in a humidity chamber. Cells were washed three times with PBS. To maintain adherence of antibodies, cells were fixed a second time with 4% PFA in PBS for 10 min and washed with PBS. Custom RNA FISH probes were designed to complement the 12S mt-RNA transcript Stellaris® RNA FISH Probe Designer (Biosearch Technologies, Inc. Petaluma,
CA) available online (www.biosearchtech.com/stellarisdesigner, version 4.2) [10](9). An adapted protocol was followed based on the manufacturer’s instructions available online (www.biosearchtech.com/stellarisprotocols). Cells were mounted onto glass slides with Pro-Long Gold Antifade Mountant (Thermo Fisher Scientific, P36930) and left to cure at room temperature for > 24 hours.

5.3.6 DNA Template For Transcription Assays

Template DNA (~500 bp) was amplified as described previously using pT7blue plasmid containing -70 to +70 of native LSP sequence [33](60) to generate ~300 bp of the transcribed downstream sequence. The reaction was carried out using PCR thermocycler with Taq 2X Master Mix (New England Biolabs) using forward U19 primer (GTTTTCCCAGTCAGTCGT) and reverse RO 300 primer (CTGGAAAGCGGGCAGTG). PCR reactions were run (95°C 2 min, 35 cycles of 95°C 15s, 55°C 30s, 68°C 30s, followed by extension of 68°C for 5 min) and were pooled together to be purified on columns from a GeneJET PCR purification kit using isopropanol precipitation to yield ~1-2 μg/μl of DNA template resuspended in 20 mM Tris-HCl, pH 7.9.

5.3.7 Protein Purification

Human TFAM was purified as previously described [10](9). Briefly, mature TFAM (res 43–246) was transformed I BL21 star (DE3) pRare E. coli. 500 ml of bacterial culture containing Dynamite media plus kanamycin and chloramphenicol was inoculated with TFAM and incubated until OD of 7. Expression was induced with IPTG at 0.5 mM, and the culture was incubated overnight at 16°C at 220 rpm. The culture was centrifuged, and the pellet was resuspended and lysed in lysis buffer [20 mM Tris-HCl, 500 mM NaCl, pH 8.0]. Cells were lysed with a microfluidizer and centrifuged for 30 minutes (70,000 g, 4°C). Protein was
purified on an immobilized metal affinity chromatography (IMAC) column and eluted in elution buffer [20 mM Tris-HCl, 500 mM NaCl, 250 mM Imidazole, pH 8.0], and dialyzed into lysis buffer [20 mM Tris-HCl, 500 mM NaCl, pH 8.0] and temporarily stored at 4°C. Nucleic acids were removed by purifying TFAM with a HiTrap Heparin High performance column (GE Healthcare). Protein was diluted to ~0.2 mg/ml in Heparin buffer (20 mM Tris-HCl, 300 mM NaCl, pH 7.5). The column was pre-equilibrated with Heparin buffer, followed by loading of diluted TFAM, and washed with 5X column volumes of Heparin buffer. Protein was eluted using a two-step gradient of 650 mM NaCl followed by 1 M NaCl. Fractions enriched in purified TFAM were pooled together and 10% (vol/vol) glycerol was added prior to flash freezing with liquid nitrogen. Proteins were stored at -80°C. Human TFB2M (res 21-396) and human POLRMT (res 120-1230) were purified as previously described [34](61).

5.3.8 Protein Labelling

Protein was fluorescently labelled using DyLight antibody labelling kits (Thermo Fisher), where TFAM was labelled with DyLight-488, TFB2M was labelled with DyLight-594, and POLRMT was labelled with DyLight-650 or -594. Fluorescent proteins were added in trace amounts to unlabeled protein (~1:100 dilutions) prior to phase separation and transcription assays.

5.3.9 Equilibrium Phase Separation Assays

Frozen protein aliquots were thawed at room temperature and kept on ice. Proteins were concentrated and/or buffer exchanged in 40 mM Tris-HCl, pH 8.0, 5 mM BME, 500 mM NaCl using 0.5 mL Centrifugal filters 10-kDa (Amicon), as needed. Protein concentration was estimated by the Bradford Assay using BSA standards (BioRad) on a Denovix spectrophotometer. To initiate phase separation under equilibrium conditions, concentrated protein
solutions at high salt (300-500 mM NaCl) and DNA (20 mM Tris-HCl, pH 7.9, 0 mM NaCl), were diluted into low-salt buffer to reach final buffer conditions of 20 mM Tris-HCl, pH 8.0, 10 mM MgCl2, 20 mM BME, 100 mM NaCl and containing either 5 or 10% PEG (MW 3,350 Da, Sigma P4338). To create the final solutions, the reagents were added to an Eppendorf tube in the following order: water and/or PEG solution that also contained trace amounts of DAPI, 10X buffer (200 mM Tris-HCl, pH 8.0, 100 mM MgCl2, 200 mM BME), 300-500 mM NaCl solution, protein (TFAM, POLRMT, and/or TFB2M), and/or DNA (see Table 1 or Figure Legend 1). Solutions were briefly centrifuged and pipetted up and down to ensure complete mixing. Approximately 3 µl of solution was added to Pluronic F-127 (Sigma, P2443) treated (for 1-3 component droplets) or untreated (for four component droplets) high-precision coverslips (Azer) and sealed with a 4.5 mm diameter × 0.6 mm depth silicon isolator (Grace-biolabs) and incubated for ~30-60 minutes at room temperature prior to imaging.
<table>
<thead>
<tr>
<th></th>
<th>TFAM (µM)</th>
<th>TFB2M (µM)</th>
<th>POLRMT (µM)</th>
<th>DNA (nM)</th>
<th>PEG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFAM+DNA</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>TFB2M+DNA</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>POLRMT+DNA</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
<td>500</td>
<td>5</td>
</tr>
<tr>
<td>TFAM+TFB2M</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>TFAM+POLRMT</td>
<td>10</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>TFB2M+POLRMT</td>
<td>0</td>
<td>10</td>
<td>1.5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>TFAM+TFB2M+DNA</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>TFAM+POLRMT+DNA</td>
<td>10</td>
<td>0</td>
<td>1.5</td>
<td>500</td>
<td>5</td>
</tr>
<tr>
<td>TFB2M+POLRMT+DNA</td>
<td>0</td>
<td>10</td>
<td>1.5</td>
<td>500</td>
<td>5</td>
</tr>
<tr>
<td>TFAM+TFB2M+POLRMT</td>
<td>10</td>
<td>1.5</td>
<td>1.5</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5.1: Compositions for mt-Nucleoid components for binary and ternary mixtures.
5.3.10 In-vitro Transcription Reaction

Reactions were set up similarly to our phase separation assays but with the incorporation of ribonucleotides (either NTPs or UTP). Briefly, water and/or PEG solution, 10X transcription buffer (200 mM Tris-HCl, pH 8.0, 100 mM MgCl2, 200 mM BME), 500 mM NaCl solution, NTPs or UTP (negative control), protein (TFAM, POLRMT, and/or TFB2M), and lastly DNA. For 1X reactions, the composition was 0.6 µM TFAM, 0.6 µM TFB2M, 0.6 µM POLRMT, and 50 nM 5 DNA. The concentration of NaCl was as follow: ~20 mM (1X), ~40 mM (3X), ~60 mM (5X), ~80 mM (7X), and ~100 mM (10X). Final concentration of buffer in reaction conditions was: 20 mM Tris-HCl, pH 8.0, 10 mM MgCl2, 20 mM BME, ~20-100 mM NaCl, 2 mM NTPs (each), 0 or 5% PEG. All other reactions were a multiple (3, 5, 7, 10X) of all reactions of the 1X reaction. Reactions were performed in a total volume of 20 µl in an Eppendorf tube on a heat block set at 35°C for one hour. Negative control for the reactions contained the same composition, except instead of 2 mM NTPs (each), 8 mM UTP was used.

Figure 5.1: Schematic diagram of transcription reactions at 35°C for 1 hour followed by fixation on glass coverslips and RNA FISH labelling (see Sub-section 5.3.14)
5.3.11 Transcription Assays Using Radionucleotides

Transcription reactions were carried out using DNA template (50 nM), mtRNAP (600 nM), TFAM (600 nM) and TFB2M (600 nM) in a transcription buffer containing 20 mM Tris (pH=7.9), 10 mM NaCl, 10 mM MgCl2, 20 mM -mercaptoethanol in the presence of ATP (0.3 mM), GTP (0.3 mM), CTP (0.3 mM), UTP (0.05 mM) and 0.1 l of [-32P] UTP (800 Ci/mmol). Where indicated, 5% PEG3350 (Sigma) was added to the transcription reaction. The concentration of all reaction components, except [α-32P] UTP, was increased proportionally in 3X, 5X, 7X, and 10X reactions. To prevent protein precipitation, NaCl concentration was increased to 35 mM (3X reaction), 55 mM (5X), 80 mM (7X), and 110 mM (10X). All reactions were incubated for 30 min at 37°C and treated with 0.2 mg/ml of proteinase K for 1 h at 55 0C. Reactions were stopped by the addition of an equal volume of 95% formamide and 0.05 M EDTA. The products were resolved by 12% PAGE containing 6 M Urea and visualized by PhosphorImager (GE Health). The amount of the transcripts produced in each reaction was quantified by ImageQuant TL software.

5.3.12 Time-course For In-vitro Transcription Reactions

Reactions were set up as before but with a total larger volume of (~60 µl). After the reaction master mix was assembled, the mix was then split into ~10 µl aliquots into five separate tubes and placed on a heat block. One tube was then removed at the indicated time points: 5, 10, 20, 40 or 60 minutes and prepared for fixation and labelling (see Sub-section 5.3.14 below).
5.3.13 Exogenous RNA Experiments

Synthetic RNA was transcribed using the SP6 promoter that was previously incorporated in the DNA template with a MAXIscript SP6/T7 transcription kit (Thermo Fisher Scientific). Synthetic RNA was purified using a RNeasy Mini Kit (Qiagen) and purified in water. Mixtures were assembled similar to the reactions as previously described, but with the addition of 8 mM UTP instead of NTPs. While setting up the mixture, RNA was added at different time points: $t = 0$ min (RNA added before protein or DNA), $t = 5$ min (RNA added after all of the components had been added and incubated on the heat block for 5 min), or $t = 30$ min (RNA added after all of the components had been added and incubated on the heat block for 30 min). In parallel, two controls were performed alongside these experiments. As a negative control, only 8 mM UTP was added, and as a positive control, NTPs (2 mM, each) were added to trigger the reaction.

5.3.14 Droplet Fixation And mt-RNA FISH Labelling

Reactions were fixed by direct addition of 16% PFA to in vitro transcription reactions in Eppendorf tubes, to create a final concentration of $\sim 0.5\%$ PFA, and the mixture was pipetted up and down several times. After $\sim 5$ minutes of incubation in the Eppendorf tube, aliquots of 3 $\mu l$ were deposited to untreated high precision glass coverslips (Azer) in 12 well plates for an additional $\sim 25$ minutes to allow fixation of condensates to the glass coverslips. Reactions were washed off with PBS, 6 leaving condensates that had become fixed to the glass coverslip and removing any soluble components in the aqueous phase. Custom RNA FISH probes (13 probes of $\sim 20$ bases) were designed to complement the mt-RNA transcript generated from the DNA template upon in vitro transcription using Stellaris® RNA FISH Probe Designer (Biosearch Technologies, Inc. Petaluma, CA) available online (www.biosearchtech.com/stellarisdesigner,
version 4.2). An adapted protocol was followed based on the manufacturer’s instructions available online (www.biosearchtech.com/stellarisprotocols). Briefly, coverslips were washed with Wash Buffer A for \(~5\) minutes. Coverslips were then treated with mt-RNA FISH probes resuspended in hybridization buffer in a humidity chamber placed in a 37°C incubator for \(~2\) hours. Coverslips were subsequently washed with Wash Buffer A for 30 minutes at 37°C. Media was aspirated and replaced with DAPI in Wash Buffer A for an additional 30 minutes at 37°C. Media was removed, and coverslips were washed in Wash Buffer B for \(~5\) minutes. All media was aspirated, and coverslips were sealed onto glass slides with Prolong Gold Antifade Mountant (Thermo Fisher Scientific) for \(>24\) hours at room temperature to allow for curing before super-resolution imaging.

**5.3.15 FRAP Experiments**

The reactions were assembled in 1.5 ml Eppendorf tubes the same as above and placed on the heat block at 35°C. After \(~30\) min, the reactions were transferred to a pre-heated 8-well imaging chamber and covered with pre-heated mineral oil (Sigma, M5904) (all at 35°C). A \(~1\) \(\mu\)m spot on the condensates was photobleached for \(~30\) s and images were acquired every 15 s.

**5.3.16 Light Microscopy**

Super-resolution imaging was performed using Structured Illumination Microscopy (SIM) on untreated droplets (equilibrium conditions), fixed droplets (after in vitro transcription), and on fixed cells at room temperature using an ELYR PS.1 on an AxioObserver Z1 inverted microscope operating ZEN software. Samples were imaged with all four laser lines (405, 488, 561, 647 nm) and imaged with a 63X/1.4 NA oil Plan Apochromat objective. Slides containing 200 nm fluorescent beads were used to standardize for channel alignment. ZEN
software was used for SIM processing and channel alignment. FRAP experiments were performed on a Carl Zeiss LSM780 microscope using a 100X objective, and droplets were imaged using transmitted light and 488/561 nm lasers. Bleaching lasted for ~30 s and images were acquired every 15s. The heating control stage insert was set to 35ºC to maintain reaction conditions.

5.3.17 Quantitative image analysis

Images were quantitatively analyzed using MATLAB, and images were visualized using ZEN, FIJI, and/or Imaris software. To characterize morphology, droplets were segmented from background using standard Matlab Image Toolbox functions and droplet features were quantified based on size, morphology, and number. To estimate degree of miscibility between pairs of components, cross-correlation coefficients were quantified for all pixels within the droplet and were averaged across all droplets per condition.

5.3.18 Bioinformatics

Protein sequences were obtained from PUBMED, and domains were annotated using Matlab [30]. Disorder was predicted from several algorithms using the Database of Disordered Protein Predictions (D2P2) [35], and graphically represented using Matlab.

5.3.19 Computational Modeling

The simulations were performed using the LaSSI simulation engine [36](44). A lattice-size, \( L \), of 120 was used for each simulation. The following move frequencies were used for all simulations:
<table>
<thead>
<tr>
<th>Move</th>
<th>Frequency</th>
<th>Normalized Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Cluster</td>
<td>0.0081</td>
<td>1</td>
</tr>
<tr>
<td>Multi-Local</td>
<td>0.0163</td>
<td>2</td>
</tr>
<tr>
<td>Double Pivot</td>
<td>0.0813</td>
<td>10</td>
</tr>
<tr>
<td>Pivot</td>
<td>0.0813</td>
<td>10</td>
</tr>
<tr>
<td>Chain Translation</td>
<td>0.0813</td>
<td>10</td>
</tr>
<tr>
<td>Snake</td>
<td>0.2439</td>
<td>30</td>
</tr>
<tr>
<td>Local</td>
<td>0.4878</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 5.2: MC move frequencies used in the simulations. Note that the last column normalizes the frequencies in the middle column such that the lowest frequency (Small Cluster) is 1.

Two-component Mixtures

For simulations of the binary mixtures, the following protocol is used. For $5 \times 10^7$ steps, the simulation temperature starts at $1000T^*$, and the following constraining potentials were applied to individual beads a distance $r$, from the center of the lattice:

$$\frac{V_1(r)}{T^*} = \begin{cases} 
  r, & r < R_c \\
  0, & \text{else}
\end{cases} \quad (5.1)$$

and

$$\frac{V_2(r)}{T^*} = \begin{cases} 
  r, & r \geq R_c \\
  0, & \text{else}
\end{cases} \quad (5.2)$$
where $V_1(r)$ is applied to Crowder chains only, and $V_2(r)$ is applied to all other components. The constraining radius,

$$
R_c = \sqrt[3]{\frac{N_{\text{molecules}} - N_{\text{crowder}}}{\pi}}
$$

(5.3)

is such that the crowder-free bead density within the bounding sphere is $\rho_0 \approx 0.75$. $N_{\text{molecules}}$ and $N_{\text{crowder}}$ correspond to the total number of beads corresponding to the interacting molecules and crowder, respectively. The constraining potential causes the non-crowder chains to be localized to the center of the lattice. The temperature is then discontinuously changed to $3.2T^*$. The system is annealed, and temperature is linearly reduced to $1.2T^*$ over $1 \times 10^8$ steps. The initial constraining potential is then turned off. The temperature is discontinuously changed to $1.1T^*$, followed by $1 \times 10^8$ steps. Again, the temperature is discontinuously reduced to $1T^*$ and $2.11 \times 10^{10}$ steps are performed. Data are acquired in the last $5 \times 10^9$ steps at a frequency of $2.5 \times 10^6$ steps. The averages from each replicate of the 5 replicates is used to generate Figure 5.14.

Table 5.3 shows the exact number of molecules used for each of the mixtures. A ratio of 1 to 5 chains is used for DNA-Protein mixtures, 1 to 2 chains for DNA-RNA mixtures, and 2 to 5 chains for RNA-Protein mixtures. This is to keep the number of modules from each species the same in the simulations.
<table>
<thead>
<tr>
<th></th>
<th>DNA</th>
<th>TFAM</th>
<th>TFB2M</th>
<th>POLRMT</th>
<th>RNA</th>
<th>Crowder</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA+TFAM</td>
<td>1000</td>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>DNA+TFB2M</td>
<td>1000</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>DNA+POLRMT</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>TFAM+TFB2M</td>
<td>0</td>
<td>5000</td>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>TFAM+POLRMT</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>TFB2M+POLRMT</td>
<td>0</td>
<td>0</td>
<td>5000</td>
<td>5000</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>RNA+DNA</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2000</td>
<td>5000</td>
</tr>
<tr>
<td>RNA+TFAM</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>2000</td>
<td>5000</td>
</tr>
<tr>
<td>RNA+TFB2M</td>
<td>0</td>
<td>0</td>
<td>5000</td>
<td>0</td>
<td>2000</td>
<td>5000</td>
</tr>
<tr>
<td>RNA+POLRMT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5000</td>
<td>2000</td>
<td>5000</td>
</tr>
</tbody>
</table>

Table 5.3: Molecule numbers for coarse-grained simulations of binary mixtures of mt-Nucleoid components.

**Full Mixtures**

For the simulations of the full set of components including RNA, the following protocol was used. For $5 \times 10^7$ steps, the simulation temperature was raised to $1000T^*$, all interactions were ignored, and the constraining potentials

$$V_3(r) = \begin{cases} r, & r < R' \\ 0, & \text{else} \end{cases} \quad (5.4)$$

and
\[
V_4(r) = \frac{T^*}{T^*} = \begin{cases} 
    r, & r \geq R' \\
    0, & \text{else}
\end{cases}
\] (5.5)

were used. \( V_3(r) \) is applied to Crowder, DNA and RNA chains, and \( V_4(r) \) is applied to TFAM, TFB2M, and POLRMT. The constraining radius

\[
R' = \sqrt[3]{\frac{N_{\text{proteins}}}{\pi}},
\] (5.6)

where \( N_{\text{proteins}} \) is the total number of beads for the proteins, is such that the bead density inside the bounding sphere is approx. 0.75. This results in TFAM, TFB2M and POLRMT localizing towards the center of the lattice. The temperature was then discontinuously reduced to \( 4T^* \), and the interactions were turned on. The constraining potentials \( V_1(r) \) and \( V_2(r) \), from above, were then applied. \( 1 \times 10^8 \) steps were performed, the system was annealed, and the temperature is linearly reduced from \( 4T^* \) to \( 2T^* \). This results in all non-crowder chains to be localized near the center of the lattice. The temperature was then discontinuously changed to \( 1.5T^* \) and \( 1 \times 10^8 \) steps were performed, followed by one more discontinuous temperature change to \( 1T^* \). \( 4.61 \times 10^{10} \) steps were performed and data were acquired at a frequency of \( 2.5 \times 10^6 \) steps over the last \( 1 \times 10^{10} \) steps, and the average was stored. The average over the 5 replicates was used to generate Figure 5.14 and Figure 5.15.

As in the binary mixture simulations, the DNA to protein ratio is kept at 1 to 5. To assess the effects of RNA, different amounts of RNA chains were included in the systems while all other components were kept fixed. RNA chains were added in increments of 100 until there were 1,000 chains, or as many RNA chains as there were DNA chains in the system. The numbers are in Table 5.4 below.
Table 5.4: Molecule numbers for coarse-grained simulations of full set of mt-Nucleoid components.

<table>
<thead>
<tr>
<th>Time</th>
<th>DNA</th>
<th>TFAM</th>
<th>TFB2M</th>
<th>POLRMT</th>
<th>RNA</th>
<th>Crowder</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>1</td>
<td>5000</td>
</tr>
<tr>
<td>$t_1$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>100</td>
<td>5000</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>200</td>
<td>5000</td>
</tr>
<tr>
<td>$t_3$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>300</td>
<td>5000</td>
</tr>
<tr>
<td>$t_4$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>400</td>
<td>5000</td>
</tr>
<tr>
<td>$t_5$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>500</td>
<td>5000</td>
</tr>
<tr>
<td>$t_6$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>600</td>
<td>5000</td>
</tr>
<tr>
<td>$t_7$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>700</td>
<td>5000</td>
</tr>
<tr>
<td>$t_8$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>800</td>
<td>5000</td>
</tr>
<tr>
<td>$t_9$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>900</td>
<td>5000</td>
</tr>
<tr>
<td>$t_{10}$</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>1000</td>
<td>5000</td>
</tr>
</tbody>
</table>

Lastly, to generate the density profiles corresponding to the components, the system center-of-mass (COM) is used as the reference point. From this reference point, radial number histograms of each component are generated. These number histograms are then normalized to generate the radial density profiles. For more details, see the supplementary index in the work of Ruff *et al.*, [37], or Sub-section 4.3.3 in Chapter 4.
5.4 Experimental Results

5.4.1 Individual Components Of The Mitochondrial Transcription Undergo Phase Separation in-vitro

Mitochondrial transcription has previously been reconstituted with three proteins and a DNA template containing a mitochondrial promoter under dilute conditions [31]. Given the organization of mt-nucleoids into condensates within the crowded mitochondrial matrix in-vivo [10], we sought to establish conditions for in-vitro mitochondrial transcription in reconstituted condensates.

Taking a bottom-up approach, we first established the individual phase behavior of the minimal components required for mitochondrial transcription. The transcription factors TFAM and TFB2M combined with the polymerase POLRMT represent the minimal components of human mitochondrial transcriptional machinery. Structural studies [38] and bioinformatics analysis show that these proteins contain a combination of ordered domains and IDRs (Figure 5.2A). Computational predictions for disordered proteins suggest that unbound TFAM is the most disordered of the three proteins with a flexible linker that bridges two DNA binding domains (High Mobility Groups A, B) and a disordered C-terminus. These features are consistent with the flexible nature and multiple conformations that have been reported for TFAM molecules in solution [39]. TFB2M and POLRMT are significantly more structured with well-folded, functional domains [30, 40], but both contain disordered regions at their N-termini. The modular nature of these proteins suggests that they have the potential to drive or contribute to phase separation of mt-nucleoids, as disordered regions contribute to non-specific, weak interactions and ordered regions provide specific, strong interactions [41].
Figure 5.2: Equilibrium phase behavior of mt-nucleoid transcriptional components. (A) Protein domain analysis for core mitochondrial transcription proteins TFAM, TFB2M, and POLRMT. Top row illustrates the known protein domains (red, blue, dark grey, respectively) and the unfolded, intrinsically disordered domains are indicated in light gray. TFAM contains two High Mobility Group domains (HMGA/B) separated by a disordered linker and flanked by a disordered C-tail. TFB2M contains an N-terminal domain (NTD) which adopts a methyltransferase fold, and the C-terminal domain (CTD), which consists of four alpha helices and a short, flexible “tail.” POLRMT is a single-subunit DNA-dependent RNA polymerase, which is distantly related to the bacteriophage T7 RNAP. It contains a hand shaped CTD that harbors the catalytic site POLRMT also contains an N-terminal domain attached to the catalytic core via a poly-proline linker. Bottom row shows probability of intrinsically disordered sequences as predicted by several models using D2P2, where high and low likelihoods for disorder are indicated in red (high) and white (low). (B) SIM images of condensates formed for individual components at room temperature on pluronic-treated coverslips: 10 µM TFAM in 10% PEG (red), 10 µM TFB2M in 10% PEG (blue), 1.5 µM POLRMT in 5% PEG (gray), and 500 nM DNA in 10% PEG (green). Bottom row contains violin plots of the aspect ratio for all condensates analyzed. n = 4 experimental replicates, average (dot) values are indicted, error bar represents the standard deviation. Scale bar = 2µm. (C) Binary and ternary compound droplets. Violin plots of correlation coefficient measured for each pair of channels; correlation coefficient = 1 denotes complete co-localization. Scale bar = 2µm. n = 4 experimental replicates, average (dot) values are indicted, and error bar represents the standard deviation. (See Table 5.1 for concentrations). (D) Quaternary droplets at room temperature. Top row includes individual channels of ∼6 µM TFAM (red), ∼6 µM TFB2M (blue), ∼6 µM POLRMT (gray), and ∼500 nM DNA (green), and the merged image in 5% PEG. Scale bar = 2µm. Bottom row contains violin plots of the correlation coefficient for all pairs of channels. n = 3 experimental replicates, average (dot) values are indicted, and error bar represents the standard deviation. Buffer for all conditions was 20 mM Tris-HCl, pH 8.0, 20 mM BME, 10 mM MgCl2, and 100 mM NaCl at room temperature.

In terms of phase behavior, we find that TFAM, TFB2M, and POLRMT as well as DNA individually demix from solution and form dense phases in the presence of the macromolecular crowder, polyethylene glycol (PEG, MW ∼3 kDa). Conditions that promote phase separation include 10 µM TFAM or TFB2M in 10% PEG; 1.5-10 µM POLRMT in 5-10% PEG; 500 nM DNA in 10% PEG (Figure 5.2B; Figure 5.3). Condensates formed by TFAM, TFB2M, or DNA resemble highly spherical droplets with an aspect ratio of ∼1 (Figure 5.2B; Figure 5.3). In
contrast, POLRMT assembles into highly irregular structures (Figure 5.2B; Figure 5.3). The differential behavior of POLRMT can be accounted for by the fact that phase separation is a density transition [42], whereby dense phases, concentrated in specific types of macromolecules, coexist with dilute phases. Density transitions can generate a spectrum of morphologies and dynamics for dense phases, especially for condensates formed by nucleic acids; moreover, the interplay between phase separation (the density transition) and percolation or gelation, which refers to the networking of macromolecules via physical crosslinks, often results in dynamically arrested phases [42–46]. Accordingly, irregular mesoscale structures are more likely to imply the formation of dynamically arrested phases, representing metastable, nonequilibrium structures, wherein one or more macromolecules are immobile because they are a part of highly crosslinked networks. In line with this interpretation, the structures formed by POLRMT show limited recoveries after photobleaching (Figure 5.3D,E) and fit the description of such dynamically arrested phases [46, 47]. Overall, our results show that all minimal components of the mitochondrial transcription machinery can undergo phase separation via homotypic interactions, yet each forms condensates with distinct dynamics and/or morphologies.
5.4.2 Multicomponent Systems Form Heterogeneously Organized And Dynamically Arrested Condensates *in-vitro*

Next, we documented the joint phase behavior of multiple components by characterizing the structures formed by binary and ternary mixtures of components of the minimal mitochondrial transcription machinery (Figure 5.2C). Using the correlation coefficient as a metric of their colocalization, we find that TFAM and DNA form multi-phase condensates, containing micron-sized sub-domains that are either TFAM-rich or DNA-rich (Figure 5.2C). In contrast, TFB2M and POLRMT show lower degrees of colocalization with DNA (Figure 5.2C). When the proteins were mixed in pairs without DNA, TFAM and TFB2M colocalize with one another although they do not colocalize as well as with POLRMT (Figure 5.2C). This again, is a likely consequence of POLRMT driving the formation of dynamically arrested phases.

For ternary combinations of pairs of proteins with DNA, the droplet organization remained inhomogeneous for all combinations, with micron-sized domains forming for TFAM-TFB2M-DNA and TFB2M-POLRMT-DNA (Figure 5.2C). However, for the ternary mixture of
all proteins without DNA, the condensates became more well-mixed, implying that the heterotypic interactions of proteins with DNA significantly contribute to the emergence of spatially distinct coexisting phases of the compound droplets (Figure 5.2C). These results demonstrate differential phase behavior of the various components of the mitochondrial transcription machinery.

Towards building transcriptionally competent condensates, all four biomolecular components (DNA, POLRMT, TFAM and TFB2M) were combined in equimolar protein ratios (∼6 μM TFAM, TFB2M, and POLRMT with 500 nM DNA) in the presence of 5% PEG (Figure 5.2D). In this mixture all biomolecules condense into single droplets (Figure 5.2D). However, these droplets do not fuse with neighboring droplets. Instead, they come into contact and form dynamically arrested higher-order structures (Figure 5.2D). This is reminiscent of arrested coalescence observed in non-biological soft matter [48]. While interfacial tension should enable fusion and a lowering of the interfacial energy, the internal network structure of the droplet provides elastic resistance to droplet deformation and fusion. Coalescence can begin, but becomes arrested when the two energies are balanced at an intermediate stage of coalescence. This arrest can happen at different stages of coalescence and is tied to the interplay of the surface energy and the elastic energy associated with the internal network structure of droplets. The extent of balancing of interfacial tension and elastic energies determines the extent of metastability and the lifetimes of the arrested phases.

We also found that localization of individual components within the multi-phasic structures remained heterogeneous: POLRMT showed the lowest level of colocalization with all other components, while TFB2M tended to accumulate more peripherally (Figure 5.2D). Importantly, the localization correlation coefficients for specific pairs tended to be higher in the compound droplet than those found in the binary or ternary droplets (Figure 5.2C-D). This behavior is consistent with the dual-nature of TFAM, which has affinity for DNA via its
N-terminal DNA-binding domain (HMGA), but also for other proteins, such as POLRMT, via its disordered C-terminus [10]. Together, these results show that the minimal mitochondrial transcription machinery can collectively form multi-phasic, dynamically arrested condensates in vitro.

5.4.3 Condensates Support Transcription And Dampen Transcription Rates

Next, we sought to test whether transcription could be reconstituted in vitro under condensate-forming conditions. In pilot experiments, we first characterized the effect of increasing concentrations of reaction components starting with standard, soluble in vitro mt-transcription reactions [31, 49]. We added a full set of nucleotides (NTPs) to the mixture of 0.6 µM TFAM, 0.6 µM TFB2M, 0.6 µM POLRMT, and 50 nM DNA in the absence of any crowder (-PEG), representing soluble conditions. After 30 minutes of incubation ~35°C (see Section 5.3), transcriptional activity was measured by detection of a ~300 nt RNA product using a PCR-amplified template containing the LSP promoter and radioactively labeled nucleotides as previously described [49] (Figure 5.4A). Transcriptional activity increased roughly linearly over a ~7-fold range of initial concentrations. This linear increase in the overall rate of transcription dropped off at higher concentrations of components (Figure 5.4A, B). High levels of TFAM led to quenching of the transcription reaction, while lower concentrations of POLRMT relative to TFAM/TFB2M reduced RNA production (Figure 5.5). By using 1:1:1 stoichiometries of proteins at approximately 10X the molar concentration of template DNA, we were able to produce significant amounts of RNA (Figure 5.4, Figure 5.5). These conditions differ from physiological conditions, in that estimates suggest a molar ratio of TFAM:mtDNA of ~1,000:1 in vivo [28], and similarly, TFAM is present in excess of TFB2M.
and POLRMT [32]. Moreover, we use a short 0.5 kb linear DNA template as longer templates require additional protein factors [38].
Figure 5.4: *In-vitro* transcription under soluble and condensed conditions leads to changes in condensate organization. (A) RNA production rates at soluble and condensed states. The transcription run-off assay was performed using the increasing concentrations (1X to 10X) of the transcription initiation complex in the absence (lanes 1-5) or presence (lanes 6-10) of PEG. Note that the apparent decrease of the transcription efficiency in lanes 5 and 10 was due, in part, to the dilution of [α-32P] UTP with “cold” UTP. (B) Quantification of RNA product from (A) as a function of reactant titration where 0% PEG is filled squares and 5% PEG is open squares. (C) Ratio of RNA production of crowded (5% PEG) to soluble (0% PEG) as a function of reactant titration. Values in A-C illustrate a representative experiment. (D) SIM images of reactions after fixation and RNA FISH, following 1 hour of reactions under the same conditions as (A). Scale bar = 5µm. All images are at the same contrast settings. (E) Quantification of RNA FISH intensity under soluble (0% PEG) or condensed conditions (5% PEG) for the reactions (NTPs) and of the negative control (UTP). n = 3 experimental replicates, values represent averages and error bars represent standard deviation. (F) SIM images of core transcription components in condensates at 7X and 5% PEG for negative control (UTP only) and for reactions (NTPs), where DNA is in green, RNA FISH is in cyan, TFAM is in red, and POLRMT is in gray scale. Scale bar = 1µm. Dashed lines indicate line profile. (G) Smoothed line profile for all components from (F).

Based on the optimization of mt-transcription under soluble conditions, we sought to reconstitute mitochondrial transcription under condensate forming conditions by proportionally increasing protein and DNA concentrations and including a crowder (+PEG). The condensates that form in the presence of the four different macromolecules (TFAM, TFB2M, POLRMT, and DNA) in the presence of PEG were transcriptionally active in the presence of nucleotides (Figure 5.4A). Interestingly, we found the transcriptional output of condensates was 1.3 - 20-fold lower than the corresponding outputs when condensates did not form (Figure 5.4B, C). Decreased rates of transcriptional output in condensates occurred most significantly first at 1X concentrations of components and approached unity with increasing reactant concentration (Figure 5.4A-C).
Figure 5.5: (A) RNA blot of transcription run-off assays showing titration of TFAM in presence (+) and absence (-) of 5% PEG using 7X concentrations: 0.35 µM DNA, 4.2 µM POLRMT, 4.2 µM TFB2M and 2-7 µM TFAM (as indicated) in transcription buffer (20 mM Tris HCl, pH 7.9, 10 mM MgCl2, 20 mM βME, 2 mM NTP (each), trace P32-GTP, and ~70-80 mM NaCl. Total RNA product relative to 2 µM TFAM-PEG are reported below the blot. Ratios of +/- PEG are reported underneath bands. (B) RNA blot of transcription run-off assays showing titration of TFAM in absence (-) of PEG for different concentrations of POLRMT using 7X concentrations: 0.35 µM DNA, 1 or 4.2 µM POLRMT, 4.2 µM TFB2M and 0.3-5 µM TFAM (as indicated) in transcription buffer (20 mM Tris HCl, pH 7.9, 10 mM MgCl2, 20 mM βME, 2 mM NTP (each), trace P32-GTP, and ~70-80 mM NaCl. Total RNA product relative to 1.2 µM TFAM for each concentration of POLRMT are reported below the blot. Ratios of 1 µM/4.2 µM POLRMT are reported underneath bands.

To confirm the formation of condensates under conditions that support transcription and to relate condensate structure to function, we compared the morphologies of condensates under
different conditions (Figure 5.4D, E; Figure 5.6). In the absence of crowder, which corresponds to the most dilute case (1X, -PEG), we did not observe condensates (Figure 5.6B-E). Under transcriptionally competent, condensate-forming conditions (1X-10X, +PEG), RNA and all transcriptional components localized to the periphery of condensates, forming vesicle-like morphologies as visualized after 60 minutes of reaction (Figure 5.4D; Figure 5.6A-D, Ref. [50]). In these vesicles, DNA tended to associate with the outermost shells, whereas RNA and proteins co-localized in the inner shell (Figure 5.6E, G). The peripheral localization of the transcription machinery appears to be a consequence of active transcription, as identical, but transcription incompetent, condensates generated in the presence of only UTP nucleotides tended to retain their filled, non-vesicular, droplet-like structures (Figure 5.4D-F; Figure 5.6F, G; see below). These results demonstrate transcriptional activity in reconstituted mitochondrial condensates and that transcription is dampened in condensates compared to in solution.
Figure 5.6: Visualizing mtRNA within in-vitro mt-nucleoid transcriptional condensates. (A-E) SIM images of reactions after RNA FISH labelling for individual channels: RNA (A), DNA (B), TFAM (C), and POLRMT (D). Top row is without PEG (-PEG) and bottom row is with PEG (5%, + PEG) for 1X-10X conditions. Scale bar = 5µm for all images. Contrast settings are equivalent for images in the same channel. (E) Images of a large vesicle after 1 hour of transcription (7X, 5% PEG, NTPs). Panels include single channel of DNA, RNA, TFAM, and POLRMT (top row), overlays of RNA/TFAM/POLRMT, DNA/RNA, DNA/TFAM, and DNA/POLRMT (bottom row), and a four-channel overlay (right). Arrows indicate the outer an inner lining of DNA. Scale bar = 1µm. (F,G) Images of condensates after 1 hour of incubation (10X, 5% PEG) with UTP (F) or NTP (G). Panels include single channel of DNA (green), RNA (cyan), TFAM (red), and TFB2M (blue). Scale bar = 1µm.

5.4.4 Newly Synthesized RNA Transcripts Shape Condensate Structures

Newly synthesized RNA transcripts localize to the periphery of mt-transcription condensates (Figure 5.4D). To determine whether nascent RNA is exclusively produced at the edge of the condensates or is generated internally and accumulates over time at the periphery, we performed time-course experiments (Figure 5.7). RNA can be detected as early as 5 minutes in the condensate interior (Figure 5.7A, B). At early time points of 5 and 10 minutes all components of the transcriptional machinery localize throughout the interior of the condensate (Figure 5.7A, B). In contrast, at intervals of 20, 40, and 60 minutes, we detected a significant change in the organization of the condensate, whereby pronounced vacuoles start to appear within the condensates, coinciding with the formation of a peripheral ring containing the RNA and transcription components and increased vacuole size with time (Figure 5.7A, B). All the components become peripherally located with increasing reaction time, and these morphological changes were concomitant with production of RNA (Figure 5.7A, B).
Figure 5.7: Dynamics of transcriptional condensate organization. (A) Time course of condensate morphology under reaction conditions (7X PEG, 5% PEG). Condensates were fixed and image after t = 5, 10, 20, 40, or 60 minutes at 35°C. Single channels for RNA (cyan), DNA (green), TFAM (red), and POLRMT (grayscale) are shown. n = 4 experimental replicates and scale bar = 1µm. Intensity of green channel was adjusted for visibility due to decrease in DAPI signal with time; all other channels are set at the same contrast settings. (B) Quantification of intensity profile of each component in the condensate, where r = 0 is the center of the condensate and r = 1 is the normalized perimeter, for each channel in (A). Shading indicates the time point corresponding to the average line profile, where darker colors are early time points and lighter colors are late time points. n = 4 experimental replicates and error bars = standard error of the mean. (C) Organization of condensates after addition of exogenous RNA (see Sub-section 5.3.13) to non-reacting droplets at t = 0 (RNA added before all other proteins/DNA), t = 5 minutes (RNA added after condensates assembled 5 min at 35°C) or t = 30 min (RNA added after condensates assembled 30 min at 35°C). Buffer was the same as that used in the negative control (8 mM UTP). Condensates were fixed onto coverslips after 1 hour of incubation at 35°C. n = 3 experimental replicates and scale bar = 0.5µm. (D) FRAP recovery for transcribing droplets (NTPs, each 2mM) for 7X and 5% PEG conditions. Inset shows condensates pre-bleach, bleach, and 9 min post-bleach. Dashed circle represents region that was bleached. Scale bar = 1µm. n = 9 droplets and error bars = standard error of the mean.

The observed reorganization of transcription components during the reaction suggests that the presence of RNA, and its increase in concentration over time, led to structural changes of the condensates. To further test the role of the newly synthesized RNA in determining condensate structure, we added exogenous RNA (ex-RNA) comparable in sequence and length to the mitochondrial components at various time points in the presence of only UTP. The goal was to mimic interactions that arise from the presence of RNA in the bulk despite the absence of transcriptional activity. Addition of ex-RNA at the beginning of mixing resulted in condensates with a layered structure, where ex-RNA, TFAM and POLRMT were in the interior, surrounded by a shell of DNA (Figure 5.7C). In contrast, addition of ex-RNA after the DNA and protein components had mixed and assembled into condensates (t = 5, 30 min) resulted in condensates that had the reverse layered organization: DNA, TFAM, and POLRMT were in the interior, surrounded by a peripheral shell of ex-RNA.
(Figure 5.7C). Here, peripheral localization of ex-RNA suggested ex-RNA coated pre-formed, existing protein-DNA-rich condensates.

Figure 5.8: FRAP for non-reacting droplets (8 mM UTP) for 7X and 5% PEG conditions. Inset shows condensates pre-bleach, bleach, and 9 min post-bleach. Dashed circle represents region that was bleached. Scale bar = 1µm. n = 10 droplets and error bars = standard error of the mean.

We found that across these conditions, TFAM and POLRMT tended to partition with ex-RNA, suggesting that these protein-RNA interactions are energetically favorable. However, in all these cases, there was little colocalization of DNA with RNA, implying that DNA and RNA repel each other and that DNA-RNA interactions are energetically unfavorable. Indeed, combining only DNA and RNA yielded condensates with layered organization, where DNA was internal surrounded by a shell of RNA, suggesting inherent immiscibility between DNA and RNA (Figure 5.8E-F). The observation that RNA is prone to forming irregular structures on its own supports that proteins mediate the miscibility between DNA and RNA in transcriptional condensates (Figure 5.8E-F).
These observations, showing that localization of components within condensates depend on the order in which different components are added, suggest that active transcription contributes to shaping condensate structure. The distinct morphologies of condensates depending on when RNA is added further indicate that active mt-transcriptional vesicular condensates represent non-equilibrium, dynamically arrested structures.

Measurements by fluorescence recovery after photobleaching (FRAP) were used to measure the mobility of actively transcribed RNA molecules by the detection of fluorescently labeled nucleotides (fluorescein-12-UTP). After \( \sim \) 30 min of reaction, minimal recovery occurred over the course of \( \sim \) 15 minutes (Figure 5.7D). As a control, free nucleotides in a transcription-incompetent condensate rapidly exchanged resulting in only limited bleach depth (Figure 5.8). Taken together, the slow internal dynamics of newly synthesized RNA molecules support our conclusion that vesicular structures formed by in vitro reconstitutions are dynamically arrested condensates.

5.4.5 Organization Of mt-Nucleoids Is Altered Upon Depletion Of Core Transcription Components \textit{in-vivo}

In intact cells, mt-nucleoids typically assemble into membraneless, nucleoprotein complexes of \( \sim 100 \) nm in diameter that are distributed throughout the mitochondrial network (Figure 5.9A) [28]. The mt-nucleoids are sites of transcription, and the newly transcribed polycistronic mtRNA localizes to adjacent mitochondrial RNA granules for further processing [11, 51]. To probe structure-function relationships of mitochondrial condensates \textit{in-vivo}, we depleted key mitochondrial transcription components and assessed their effects on mt-nucleoid fine-structure.
Figure 5.9: Structural changes upon perturbation of mt-nucleoid components in live cells. (A,B) SIM images of mitochondrial components after 72 hours of siRNA treatment: siNEGATIVE (A) and siTFAM (B). Leftmost panels are the merged image of the zoomed-out version of the mitochondrial network, where mitochondria are in magenta (MitoTracker Red), mtDNA is in green (anti-DNA), TFAM is in red (anti-TFAM), and 12S mt-rRNA is in cyan (RNA FISH). Scale bar = 1 \( \mu m \). White box indicates region of interest (ROI). Middle panels are signal or two-channel overlays of the ROI. Rightmost panels are the three-channel overlay for the ROI. Scale bar = 0.2 \( \mu m \). Intensities across each channel are matched across siRNA conditions. \( n = 4 \) independent experimental replicates. (C) Probability distribution of TFAM pixel intensity within a segmented nucleoid, where gray = siNEGATIVE and magenta = siTFAM. Nucleoids containing higher than normal TFAM intensities are considered to be stressed as indicated by black arrow. (D) Probability distribution of DNA pixel intensity within a segmented nucleoid, where gray = siNEGATIVE and magenta = siTFAM. (E) RNA/TFAM intensity for stressed nucleoids (see arrow in D). Data represent \( n = 4 \) independent experimental replicates that were pooled together.
While complete knockdown of TFAM is embryonically lethal, partial knockdown of TFAM by RNAi (Figure 5.8; Figure 5.10; Figure 5.11) leads to a significant reorganization of mt-nucleoids as previously observed in TFAM heterozygous knockout mice and analogous RNAi experiments [52, 53]. The number of mt-nucleoids per cell, based on staining for TFAM and DNA, was reduced significantly (Figure 5.10). In line with prior observations of induction of a stress response involving increased interferon stimulated gene expression and enhanced type I interferon responses upon loss of TFAM [52], dramatically enlarged clusters of mt-nucleoids were observed in TFAM depleted HeLa cells (Figure 5.9B). Interestingly, these remaining stress-induced mt-nucleoids resembled the heterogeneous condensates observed in-vitro [10], where purified TFAM and mtDNA form > 1 micron-sized, multiphase droplets. Indeed, these enlarged mt-nucleoids allowed us resolve the spatial organization of the mitochondrial (ribo)nucleoprotein complexes (Figure 5.9B): mitochondrial RNA localized peripherally, demixed from TFAM and mtDNA (Figure 5.9B, Figure 5.11), similar to in-vitro condensates we formed when ex-RNA was added at later time points (Fig. 3C), and supporting the conclusion that mt-nucleoids and mtRNA exist as spatially distinct phases in live mitochondria [11, 14].
Figure 5.10: High throughput confocal images after siRNA knockdown. (A) Maximum intensity images from high-throughput confocal imaging after siRNA knockdown of most significant conditions, where left panels are of 12S rRNA (RNA FISH, cyan), middle panels are of TFAM (anti-TFAM, gray scale), and right panels are of mitochondria (MitoTracker Red, magenta), for untreated, siNEGATIVE, siTFAM, and siMTERF2. Images shown represent approximately 25% of entire field of view as detected by camera on CV7000 Yokogawa high-throughput cytological discovery system. Scale bar = 50 µm. (B) Quantitative analysis of high-throughput imaging after siRNA treatment, including number of mt-nucleoids per cell, TFAM brightness per mt-nucleoid, and number of cells. Analysis is of one experimental replicate with three technical replicates, each with five fields of view, where error bars = standard deviation. $P$-value for ANOVA test was $P < 0.01$ for number mt-nucleoids per cell, $P < 0.01$ for TFAM brightness per mt-nucleoid, and $P < 0.001$ for cell number. For individual pairs, a least significant difference procedure was performed, where $^* P < 0.05$, $^{**} P < 0.01$, and $^{***} P < 0.001$. (C) Left: Western blot confirming TFAM (anti-TFAM) depletion for untreated, siNEGATIVE, and siTFAM in HeLa cells after 72 hours. Loading control was actin (anti-beta-actin). Right: Western blot confirming mTERF2 (anti-mTERF2) depletion for untreated, siNEGATIVE, and siMTERF2 in HeLa cells after 72 hours. Loading control was tubulin (anti-tubulin-alpha).
We further noticed altered phase behavior upon perturbation of other mt-nucleoid components involved in transcription (Figure 5.10). The mt-nucleoid-associated protein MTERF2 (mitochondrial transcription termination factor 2), is an abundant mt-nucleoid protein, present at \( \sim 100:1 \) copies relative to mtDNA \textit{in-vivo} [54]. Partial knockdown of MTERF2 also leads to a stress response with significantly reduced cell number and altered mt-nucleoid number (Figure 5.10), associated with a pronounced population of swollen mitochondria. These mitochondria corresponded with ring-like accumulation of RNA puncta alongside the rounded membrane (Figure 5.11A), reminiscent of the RNA peripheral localization of active \textit{in-vitro} transcription condensates (Figure 5.4; Figure 5.7). TFAM accumulated in these mitochondria and appeared to wet the inner surface of 12S RNA foci. However, mtDNA remained organized as \( \sim 100 \) nm puncta surrounded by TFAM that were frequently positioned adjacent to bright puncta of 12S RNA (Figure 5.11), which further supports the idea of coexisting (ribo)nucleo-protein phases. In contrast, depletion of mtDNA achieved using a mitochondrially targeted endonuclease [55], led to complete dissolution of mt-nucleoids, including TFAM, and reduction of 12S RNA signal (Figure 5.11). The dissolution of DNA- and RNA-rich condensates in the mitochondrial matrix upon mtDNA depletion suggests that mtDNA nucleates mitochondrial transcriptional condensates in live cells.
Taken together, our results support the notion that the (ribo)nucleoprotein complexes in the mitochondrial matrix represent coexisting phases, whose organization is modulated by the physicochemical nature of the constituents as well as functional activity of the mt-nucleoid, specifically RNA production during active transcription.

5.5 Computational Results

To synthesize all our findings into a model that describes the collective phase behavior of active transcriptional condensates, we used computational modeling to recapitulate the effect of RNA production on phase behavior of mt-nucleoids. Using the simulation engine LaSSI [36], we performed Monte-Carlo (MC) simulations of coarse-grained (CG) models of binary and higher-order mixtures to probe the effects of RNA production on equilibrium condensates (Fig. 5). To preserve the overall length-scales and interaction hierarchies of the macromolecules involved in transcriptionally active condensates in the simulations, DNA
molecules were modeled as chains of twenty beads, RNA as chains of ten beads, TFAM and TFB2M as chains of four beads, POLRMT as chains of three beads, and crowders as chains of four beads (Figure 5.12).

Figure 5.12: Coarse-grained model for the mitochondrial transcriptional components. All beads are connected by implicit linkers of 2-lattice sites. DNA is modeled as chains of 20 beads. Since the mtDNA template used in the experiments produces RNA about half as long as the DNA, the RNA is a chain of 10 beads. TFAM is modeled as 4 beads (‘X-a-b-X’) where the central two beads, denoted as beads a and b, interact more favorably with DNA. To account for TFAM’s weak dimerization upon DNA binding, bead b has an additional interaction energy ($-2k_B T$) for another b bead, yielding a local anisotropic interaction. TFB2M is represented as 4 beads, and POLRMT is represented as 3 beads due to the highly folded nature of the molecule. Lastly, the Crowder is represented as a chain of 4 beads.

5.5.1 Matching Experimental Phase Behavior With Coarse-Grained Model

We first parameterized the contact energies between pairs of molecules by reproducing experimentally observed morphologies of single and binary mixtures (Figure 5.14, compare to Figure 5.2B, C) [10]. The experimentally measured colocalization and condensate homogeneity were used as proxies for stronger heterotypic interactions, while spatial inhomogeneities within
the condensates were used as proxies for stronger homotypic interactions. TFAM has an additional anisotropic interaction to account for its weak dimerization upon DNA binding (Figure 5.13) [56].

Figure 5.13: Interaction matrices for the two models considered. Model A lacks RNA-binding proteins, and recaptures the organization of the condensates seen in vitro. The Crowder has repulsive interactions with every species, including itself. The DNA and RNA have no interactions, while the rest of the molecules have favorable interactions. Model B mimics the effective inclusion of RNA-binding proteins by making the RNA-Crowder interactions to be favorable, modeling an effective RNA binding protein.

Simulations of the binary mixtures of transcriptional condensate components generated morphologies (Figure 5.14; Figure 5.15) with spatial organizations, quantified in terms of radial density profiles, that recapitulate the experimental results (Figure 5.2C). For example, TFAM-DNA and POLRMT-DNA formed multi-phase droplets in the simulations, while TFB2M and DNA behaved as distinct coexisting phases, and pairs of proteins tended to form well-mixed droplets, as observed in-vitro (Figure 5.2C, Figure 5.14).
Figure 5.14: Radial density profiles for main pairs of mt-nucleoid transcriptional machinery. Representative snapshots and density profiles for the binary mixtures shown in Figure 5.2. For clarity, the Crowder is not shown. Model A qualitatively captures the experimental images. DNA & TFAM form mixed condensates with TFAM having smaller droplets within. DNA & TFB2M generate droplets that do not fully colocalize but stay closely associated. TFAM & TFB2M form well mixed droplets. POLRMT mixtures have dense POLRMT regions covered by the other component.
Figure 5.15: Radial density profiles of mt-nucleoid transcriptional machinery with RNA. Representative snapshots and density profiles for all binary mixtures containing RNA with Model A. For clarity, the Crowder is not shown. RNA has favorable interactions with TFAM, TFB2M and POLRMT and thus generates well mixed droplets in those cases. Since RNA and DNA do not have favorable interactions, RNA and DNA make separate droplets.

Extending this modeling approach to the quaternary condensates, we find the formation of heterogeneous droplets that are consistent with our experimental results (Figure 5.18, \( t = 0 \); compare to Figure 5.2D). Simulations show the formation of layered droplets with DNA being localized almost exclusively at the periphery (Figure 5.16; Figure 5.17). The three proteins tend to be enriched in the interior of the droplet, enveloped by a shell of DNA. This multi-phase organization is consistent with the heterogeneity and co-localization of components in the four-component dynamically arrested droplets observed \textit{in-vitro} (Figure 5.2D).
Figure 5.16: Radial density profiles as a function of time for individual components of mt-nucleoid transcriptional machinery (Model A). Density profiles of each component with different RNA amounts, corresponding to different times. DNA is pushed outwards as RNA is increased. For TFAM the interior most density does not change but the density at the outer rim shared with DNA is reduced. TFB2M is significantly pushed outwards as RNA is increased. RNA is accumulated evenly inside the interior of the droplet as RNA is increased. The inset corresponds to Model B and shows that no RNA is accumulated inside the condensate when simulated RNA binding proteins are present.
5.5.2 Condensate Reorganization via RNA Transcription

Based on our experimental observations of reorganization of transcriptional components upon synthesis of nascent RNA in the condensate, we sought to explore how transcription affects the morphologies of condensates \textit{in-silico}. At the initial time point $t_0$, all reactants are fully mixed, and no RNA is present, while at some later time, $t_{i>0}$, the system contains RNA, $[R]_{i>0}$; conversely, if the system has an amount of RNA, $[R]_i$, then the system is at $t_i$. Therefore, titrating the amount of RNA in the simulations allowed us to model the equilibrium structures of condensates that should form when transcription occurs (Fig. 5). We find that
owing to the favorable interactions between the RNA and the proteins, newly synthesized RNA is readily incorporated into the existing droplet (Figure 5.18). RNA distributes itself into the interior of the droplets (Figure 5.16; Figure 5.17), and with increasing levels of RNA, TFB2M is peripheralized from the interior, which further pushes DNA to be localized almost exclusively to the periphery (Figure 5.16). Therefore, in accord with experimental observations, the simulations show that transcription directly remolds the organization of the condensate components. Strikingly, the computationally predicted equilibrium structures are exact facsimiles of structures we obtain upon mixing ex-RNA with components early on, \( t = 0 \) (Figure 5.7C).

![Figure 5.18](image)

Figure 5.18: Representative snapshots of simulated droplets of Model A as a function of time. A higher amount of RNA in the system corresponds to a later time in the transcriptional reactions. For clarity, the Crowder is not shown, and non-RNA components are made transparent. The RNA evenly distributes itself inside the condensates and continues to be accumulated as the reaction continues.

### 5.5.3 RNA-binding Proteins Can Prevent Reorganization Due To RNA Transcription

While we find that RNA accumulates in vesicular condensates upon transcription \( \textit{in-vitro} \), this organization is not observed \( \textit{in-vivo} \). It is likely that in non-stressed cells newly synthesized RNA molecules are bound by RNA-binding proteins and processed in mitochondrial RNA
granules, which are phase separated structures that are often located adjacent to mt-nucleoids [11, 51]. To test whether the action of RNA-binding proteins may account for the distinct RNA distribution in DNA-rich condensates *in-vitro* and *in-vivo*, we incorporated an additional favorable interaction between the crowder and the RNA in our simulations to mimic the effects of association of RNA-binding proteins to the newly synthesized RNA under steady-state conditions *in-vivo* (Figure 5.13).

![Figure 5.19: Representative snapshots of simulated droplets of Model B as a function of time. A higher amount of RNA in the system corresponds to a later time in the transcriptional reactions. For clarity, the Crowder is not shown, and non-RNA components are made transparent. With a suitable RNA binding protein, the RNA can be prevented from going inside the condensates (Figure 5.18).](image)

Under these conditions, the reorganization of transcriptional components mirrors their distribution *in-vivo* (Figure 5.9; Figure 5.10; Figure 5.11), where RNA no longer accumulates within the DNA-rich droplet but condenses separately in the bulk (inset Figure 5.16; Figure 5.18; Figure 5.20). Overall, these simulations recapitulate the *in-vivo* behavior of mitochondrial condensates, and they also affirm the inference that vesicular structures observed *in-vitro* are non-equilibrium, dynamically arrested phases.
Figure 5.20: Radial density profiles as a function of time for individual components of mt-nucleoid transcriptional machinery (Model B). Density profiles of each component, in separate panel, with Model B, as a function of RNA amount, or reaction time. Since the RNA can favorably interact with the Crowder, we can see that there is no RNA inside the transcriptional condensate. TFB2M and POLRMT, which are mostly towards the interior are unaffected by the increased amount of RNA. The outermost layers which include DNA and TFAM are slightly affected. As more RNA is included, more of the Crowder is engaged in interactions with RNA, which reduces the effective amount of Crowder for the transcriptional components.

5.6 Discussion

We report here the reconstitution of transcriptionally active, multi-phasic condensates using the human mitochondrial transcription machinery as a model system. We show that, when compared to bulk reactions realized in the absence of condensates, the transcriptional rate is reduced under condensate-forming conditions. This, we attribute to the dynamically arrested nature of transcriptionally active condensates that form in-vitro. We also observe that the production of RNA alters the spatial organization of condensates, thus providing direct evidence for a dynamic interplay between the structure of condensates and the functional activities they harbor.

Decreased rates of transcription within condensates that form in-vitro are likely associated with slower internal dynamics that we observe in these condensates. These may be thought of as increasing the Damköhler numbers \( D_a \sim k/D \), which compare reaction rates \( k \) relative
to the diffusive mass transfer rates (diffusion coefficient, $D$) [57]. Whereas in bulk solution, reactants can diffuse quickly, and the rate of transcription is only limited by the speed of the polymerase ($D_a \ll 1$, reaction-limited), the significantly slow dynamics associated with the condensate environment, as they exist in mitochondrial condensates *in-vivo*, suggest that condensates experience diffusion- or transport-limited kinetics ($D_a \gg 1$, diffusion-limited) [12]. Our rough estimates suggest that less than one round of transcription occurs under condensate-forming conditions indicating that not all DNA templates were actively transcribed. This behavior is in line with the situation in live cells, where only a minority of $< 5\%$ of the mitochondrial nucleoids are actively transcribed at any given time [58, 59]. Comparisons of *in-vitro* and *in-vivo* mitochondrial transcription properties and phase behavior are complicated by difference in the stoichiometry and DNA templates required to reconstitute efficient mt-transcription *in-vitro* under both soluble and condensate conditions. For example, *in-vivo* TFAM is present in roughly 1,000 copies per copy of mtDNA [28] and in significant stoichiometric excess of TFB2M and POLRMT [32], whereas *in-vitro* reconstitution systems require use of roughly equimolar ratios of proteins to generate detectable RNA product as relatively high TFAM levels significantly quench the reaction (Figure 5.5) [31]. Moreover, the mitochondrial genome is circular and 16 kb in size, whereas *in-vitro* transcription assays use short linear DNA templates of $\sim 0.5$ kb, since longer transcripts cannot be efficiently generated *in-vitro* without the presence of other protein factors [38]. The RNA molecules generated *in-vivo* are long transcripts that are quickly bound and modified by RNA-binding proteins [60], whereas our *in-vitro* system generates short ($\sim 300$ nt) RNAs in the absence of any such RNA-modifying proteins.

There are growing reports of transcription occurring within condensates *in-vivo*. RNA Pol I transcribes rRNA in the multiphase nucleolus [8, 61], RNA Pol II has been reported to produce mRNA in transcriptional condensates [62, 63], and POLRMT generates long,
polycistronic mtRNA in mitochondrial condensates [11, 14, 38]. While condensate formation is not an absolute requirement for transcription [33], there is growing evidence that condensate formation may offer unique advantages for regulation of transcription in-vivo. First, the condensed phase enriches for specific reactants, which may regulate mass action effects [12]. For example, SUMOylation can be increased ∼36 fold in engineered condensates in-vitro [34]. In addition, increased dwell times of proteins associated with the condensate microenvironment may also be conducive for assembly of reacting complexes [35]. In support of this hypothesis, an early FRAP study showed that the kinetic behavior of multiple RNA Pol I components could only be explained by inclusion of a slow kinetic component prior to binding of the polymerase subunits to the promoter, possibly reflecting slowed diffusion in a nucleolar condensate [64].

Importantly, we find that the structure of the transcriptional condensate is affected by its activity. Production of a new chemical species-RNA-in the otherwise DNA- and protein-rich transcription condensates leads to non-equilibrium changes in condensate organization, reflected by the emergence of vesicle-like morphologies. Intriguingly, similar vesicle morphologies have been observed for simple in-vitro RNA-protein systems [50], in-vivo liquid spherical shells of the DNA- and RNA-binding protein TDP-43 [65], and in-vitro liquid spherical shells of DNA and poly-L-Lysine [66]. RNA has also been shown to form a corona on the surface of engineered condensates, directly impeding their coarsening [67]. This implies that vesicular structures are likely to be thermodynamically accessible in specific regions of protein-nucleic acid phase diagrams. Importantly, while equilibrium simulations predict the organization of multiphase condensates at the onset of the transcription reaction, they do not produce vesicle-like structures as transcription progresses. This reinforces our conclusion that the vesicles we observe in-vitro are non-equilibrium structures.
The vesicle formation observed due to RNA generation in transcriptionally active condensates in-vitro differs from that of the canonical mt-nucleoid organization in mitochondria in-vivo [28, 29], where mt-nucleoids remain as condensed, ~100 nm droplet-like structures, and mtRNA localizes to separate granules that coat the mitochondrial inner membrane and are de-mixed from the mt-nucleoid [11]. Our simulations, which are based on experimentally determined interaction parameters, suggest that during transcription, mtRNA is tethered to the mt-nucleoid as it is being transcribed by POLRMT but later moves - be it actively or passively - out of the mt-nucleoid into more energetically favorable RNA processing granules in mitochondria. An attractive model, consistent with our observations, is that in-vivo the nascent mtRNA is effectively handed off from the mt-nucleoids condensates, from where it is generated, to the more energetically favorable mitochondrial RNA processing granules, which are themselves condensates [11]. This scenario is supported by our computational simulations which demonstrate that the presence of RNA-binding activities outside the condensate is sufficient to remove the accumulating RNA from the condensate.

Finally, our observations regarding mt-transcription condensates are also relevant to other cellular transcriptional processes. Transcription of ribosomal DNA (rDNA) by RNA Pol I occurs in fibrillar centers (FC) in the nucleolus, which unlike the mt-nucleoid, contains enveloping coexisting phases of RNA and proteins [8]. However, both mt-nucleoids and FCs represent cores of a DNA-rich phase (mtDNA or rDNA, respectively), in which the nascently transcribed RNA moves radially outwards, into adjacent mtRNA granules in mitochondria [51] or towards the dense fibrillar component (DFC) and granular components (GC) in the nucleolus [61], respectively. Similarly, RNA Pol II can self-assemble with the transcriptional machinery to form condensates, particularly at super-enhancers, in the mammalian nucleus [62]. Interestingly, for RNA Pol II condensates, rapid, local RNA production has been shown to result in complete dissolution of the condensate, underscoring a feedback mechanism between
phase behavior and RNA production as also observed here for mitochondrial transcription [26]. Similar condensation is also seen in the bacterial nucleoid, where RNAP clusters with transcription factors at specific sites, particularly rDNA, in the bacterial genome [68, 69]. The common observation that emerges from these observations is that RNA is not retained in DNA-rich phases, pointing to an intrinsic energetic barrier for their mixing.
5.7 References


53. Aasumets, K., Basikhina, Y., Pohjoisnäki, J. L., Goffart, S. & Gerhold, J. TFAM knockdown-triggered mtDNA-nucleoid aggregation and a decrease in mtDNA copy

239


Chapter 6

RNA Binding Proteins Form Mesoscale Clusters In Subsaturated Solutions

6.1 Preamble

6.2 Introduction

Phase separation of RNA binding proteins with disordered prion-like domains (PLDs) and RNA binding domains (RBDs) is implicated in the formation and dissolution of membraneless biomolecular condensates such as RNA-protein (RNP) granules [2–10]. Macroscopic phase separation is a process whereby a macromolecule in a solvent separates into a dilute, macromolecule-deficient phase that coexists with a dense, macromolecule-rich phase [11, 12]. In a binary mixture, the soluble phase, comprising dispersed macromolecules that are well-mixed with the solvent, becomes saturated at a concentration designated as $c_{\text{sat}}$. Above $c_{\text{sat}}$, for bulk concentrations $c_{\text{tot}}$ that are between the binodal and spinodal, phase separation of full-length RNA binding proteins and PLDs is thought to follow classical nucleation theory [13–16].

In classical nucleation theories, clusters representing incipient forms of the new dense phase form within dispersed phases of supersaturated solutions defined by $c_{\text{tot}} > c_{\text{sat}}$ [17, 18]. In the simplest formulation of classical nucleation theory [17–19], the free energy of forming a cluster of radius $r$ is:

$$
\Delta F = -\frac{4\pi}{3} r^3 \rho_n \Delta \mu + 4\pi r^2 \gamma 
$$

(6.1)

Here, $\Delta \mu$ is the difference in the chemical potential between the one-phase and two-phase regimes, which is negative in supersaturated solutions and positive in subsaturated solutions;
$\rho_n$ is the number of molecules per unit volume and $\gamma$ is the interfacial tension between dense and dilute phases. At temperature $T$, in a seed-free solution, the degree of supersaturation $s$ is defined as:

$$s \equiv \frac{\Delta \mu}{RT} = \ln \left( \frac{c_{\text{tot}}}{c_{\text{sat}}} \right)$$

(6.2)

where $R$ is the ideal gas constant. Here, $s$ is positive for $c_{\text{tot}} > c_{\text{sat}}$, and as $s$ increases, cluster formation becomes more favorable. Above a critical radius, $r^*$, the free energy of cluster formation overcomes the interfacial penalty, and the new dense phase grows in a thermodynamically downhill fashion. Ideas from classical nucleation theory have been applied to analyze and interpret the dynamics of phase separation in supersaturated solutions [13, 14, 16]. Classical nucleation theories stand in contrast to two-step nucleation theories that predict the existence of pre-nucleation clusters in supersaturated solutions [20–23]. These newer theories hint at the prospect of there being interesting features in subsaturated solutions i.e., for $c_{\text{tot}} < c_{\text{sat}}$ and $s < 0$.

Recasting Equation (6.1) in terms of the number of molecules per cluster, $n$, gives us:

$$\Delta F = -n\Delta \mu + n^{2/3}\gamma',$$

(6.3)

where $\gamma' \equiv (36\pi)^{1/3}v^{2/3}\gamma$ is simply a rescaled interfacial tension, and where $v$ is the volume occupied by a single molecule. The subsaturated regime, where $s$ is negative, corresponds to the one-phase regime. Above $c_{\text{sat}}$, the dynamics of phase separation can be described by classical nucleation theories. Ignoring the interfacial tension, $\gamma' = 0$, the free energy of realizing clusters with $n$ molecules becomes: $\Delta F = -n\Delta \mu$. The probability $P(n)$ of forming a cluster of $n$ molecules in a subsaturated solution is proportional to $\exp(sn)$. Accordingly, the relative probability $P(n)/P(1)$ of forming clusters with $n$ molecules will be $\exp(s(n - 1))$. 244
The relative probability of forming clusters with \( n \) molecules, which may be thought of as the concentration of clusters with \( n \) molecules, is negligibly small for clusters with more than a few molecules. This is true irrespective of the degree of subsaturation. Is this expectation valid? We show here that subsaturated solutions feature a rich distribution of species not anticipated by classical nucleation theories. Here, we report results from measurements of cluster size distributions in subsaturated solutions of phase-separating RNA binding proteins from the FUS-EWSR1-TAF15 (FET) family. We find that these systems form heavy-tailed distributions of clusters in subsaturated solutions. The abundant species are always small clusters. However, as bulk concentrations \((c_{\text{tot}})\) increase, the distributions of cluster sizes shift continuously toward larger values. We discuss these findings in the context of theories for associative polymers [10, 24–31].

### 6.3 Materials And Methods

#### 6.3.1 Materials

The materials used are listed in Table 6.1 below.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemicals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIS</td>
<td>Carl Roth Germany</td>
<td>77-86-1</td>
</tr>
<tr>
<td>Potassium Chloride (KCl)</td>
<td>Merck Germany</td>
<td>7447-40-7</td>
</tr>
<tr>
<td>Reagents</td>
<td>Source</td>
<td>Identifier</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Chemicals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>VWR chemicals</td>
<td>56-81-5</td>
</tr>
<tr>
<td>cOmplete™</td>
<td>Roche Germany</td>
<td>11697498001</td>
</tr>
<tr>
<td>Imidazole</td>
<td>Sigma Aldrich Germany</td>
<td>288-32-4</td>
</tr>
<tr>
<td>DTT</td>
<td>Alfa Aesar Germany</td>
<td>578-51-7</td>
</tr>
<tr>
<td>Hexane di-ol</td>
<td>Sigma Aldrich Germany</td>
<td>629-11-8</td>
</tr>
<tr>
<td>ATP-Na</td>
<td>Jena Bioscience</td>
<td>987-65-5</td>
</tr>
<tr>
<td>Maltose</td>
<td>Sigma Aldrich</td>
<td>6363-53-7</td>
</tr>
<tr>
<td>Bacterial and Virus Strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sf9 cells</td>
<td>Expression Systems</td>
<td>Cat#94-001F</td>
</tr>
<tr>
<td>Recombinant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUS GFP</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH1204</td>
</tr>
<tr>
<td>FUS SNAP</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH0901</td>
</tr>
<tr>
<td>FUS</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH0901</td>
</tr>
<tr>
<td>FUS PLD (1-211)</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH0994</td>
</tr>
<tr>
<td>FUS (RBD, 10D/4E-G)</td>
<td>This study</td>
<td>TH1740</td>
</tr>
<tr>
<td>FUS (PLD,10Y-S)</td>
<td>This study</td>
<td>TH1427</td>
</tr>
<tr>
<td>FUS (PLD, 18Y-S)</td>
<td>This study</td>
<td>TH1815</td>
</tr>
<tr>
<td>FUS (RBD, 6F-G)</td>
<td>This study</td>
<td>TH1917</td>
</tr>
<tr>
<td>FUS (RBD, 6Y-S)</td>
<td>This study</td>
<td>TH1918</td>
</tr>
<tr>
<td>TAF15 SNAP</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH1203</td>
</tr>
<tr>
<td>Reagents</td>
<td>Source</td>
<td>Identifier</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Chemicals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EWSR1 SNAP</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH1276</td>
</tr>
<tr>
<td>hnRNPA3</td>
<td>Wang et. al. 2018 [10]</td>
<td>TH1202</td>
</tr>
</tbody>
</table>

### 6.3.2 Constructs, Protein Expression And Purification: Method A

The construct / protein sequences used are listed in Section C. Details of materials used for the preparation of samples are as follows:

- **Lysis buffer**: 50 mM Tris-HCl pH 7.4, 1 M KCl, and 5% Glycerol.

- **Protease inhibitor**: cOmplete™, EDTA-free Protease Inhibitor Cocktail Tablets.

- **NTA elution buffer**: 50 mM Tris-HCl pH 7.4, 1 M KCl, 5% Glycerol and 300 mM Imidazole.

- **MPB elution buffer**: 50 mM Tris-HCl pH 7.4, 1 M KCl, 5% Glycerol and 30 mM Maltose. **Storage buffer** 50 mM Tris-HCl pH 7.4, 500 mM KCl, and 5% Glycerol, and 1 mM DTT.

### Purification Method

All proteins were first expressed in 1 L SF9 (1 million/mL) insect cells with 5 mL P2 virus and harvested 72 hours post infection. Cells were collected by centrifugation for 5 min at 2,000 rpm. The pellets were re-suspended in lysis buffer. Protease inhibitors were added to the lysis buffer solution. The cells were lysed by sonication. The crude lysate was clarified by centrifugation for 20 min at 17,000 rpm. After centrifugation, the supernatant was passed
through with Ni-NTA agarose column using Ismatec peristaltic pump. The protein-bound beads inside the column were further washed with 10 column volumes of lysis buffer and the mixture of 10 mM of Imidazole with lysis buffer, respectively, to remove non-specific bound proteins from the Ni-NTA column. The proteins were eluted with the NTA elution buffer.

The eluted protein solution was passed through MBP resins using gravimetric column. The protein bound beads were washed with 10 column volume of lysis buffer. The proteins were eluted with the MBP elution buffer. The entire process was monitor using Bradford assay solution and each fraction were investigated with Gel electrophoreses.

To cleave the His-MBP tag, 3C prescission protease was added to the eluted protein at a 1:100 ratio. The mixture was incubated at RT for 4 hours and was purified over the gel filtration chromatography (ÄKTA with Superdex-200 increase column; GE Healthcare) equilibrated with storage buffer. Peak fractions were pooled and immediately use for the DLS or NTA experiments. Further, to obtain untagged protein TEV protease were added at a 1:50 ratio and incubated at RT for 6 hours. The untagged protein was purified over the gel filtration chromatography (ÄKTA with Superdex-200 increase 10/300 column; GE Healthcare) equilibrated with storage buffer. Peak fractions were pooled and concentrated using Amicon 15 30.000 MWCO at 4000 rpm in RT. Protein concentration was determined by measuring absorbance at 280 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The 260/280 ratio of all purified proteins were measured between 0.52 to 0.56.

(Note: Proteins were stored with tags at -80 C, and prior to experiments the proteins were thawed, tag was cleaved, run gel filtration chromatography, peak fractions were pooled and concentrated and used for the experiments. This way the data reproducibility was good. If the proteins were frozen and thawed, the reproducibility of the experiments varies.)
6.3.3 Constructs, Protein Expression And Purification: Method B

FUS SNAP protein expression and purification using method B. Details of materials used for the preparation of samples are as follows:

- **Lysis buffer:** 50 mM Tris/HCl, pH 8.2, 1 M NaCl, 25 mM Imidazole, 5% (w/v) Glycerol, 1 mM DTT.

- **Buffer A:** 50 mM Tris/HCl, pH 8.2, 1 M NaCl, 25 mM Imidazole, 5% (w/v) Glycerol, 1 mM DTT.

- **Buffer B:** 50 mM Tris/HCl, pH 8.7, 1 M NaCl, 250 mM Imidazole, 5% (w/v) Glycerol, 1 mM DTT.

- **Buffer C:** 50 mM Tris/HCl, pH 8.7, 1 M NaCl, 250 mM Imidazole, 5% (w/v) Glycerol, 1 mM DTT, 10 mM Maltose.

- **Buffer D:** 50 mM Tris/HCl, pH 7.5, 0.5 M KCl, 5% (w/v) Glycerol, 1 mM DTT.

**Purification Method**

FUS-SNAP was expressed in 2 L SF9 (2 million/mL) insect cells with 20 mL P3 virus and harvested 72 hours post infection. Cells were collected by centrifugation for 5 min at 1,500 rpm. The pellets were re-suspended in lysis buffer. Protease inhibitors were added to the lysis buffer solution with 100 µL of Benzonase. The cells were lysed by using a shear fluid homogenizer (LM10 Microfluidizer from Microfluidics) at 5000 psi. The crude lysate was clarified by centrifugation for 60 min at 25,000 rpm at 15°C.

3 × 5mL HisTrap Cytiva Columns Position 2, 25 mL HiPrep Amylose (NEB HiFow Amylose 25 ml XK16/20) Position 8, and 320 mL Superdex 200 pg 26/60 Column Position 7 were
mounted in ÄKTA explorer. HisTrap Columns were equilibrated with Buffer A, loading cell lysate followed by washing with Buffer A and elution with Buffer B (3 column volume). HiPrep Amylose column was equilibrated with Buffer B, loading the eluted protein from HisTrap Columns, followed by washing with Buffer B. 3C PreScission Protease was loaded into the Amylose column and incubated for 8 hours at RT. The protein was eluted using Buffer B. The Amylose column was regenerated using Buffer C. Superdex 200 pg column was equilibrated with Buffer D. The eluted protein from Amylose column was loaded and run the SEC using Buffer D. Peak fractions were pooled and concentrated using Amicon 15 10.000 MWCO at 15000 rpm in RT. Protein concentration was determined by measuring absorbance at 280 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The 260/280 ratio of all purified proteins were measured between 0.55. The concentrated protein was aliquoted, and flash frozen with liquid N2.

6.3.4 Phase Separation Assays

For droplet formation in the absence of crowding agents, proteins were diluted into various concentrations in the corresponding buffers in a total solution volume of 25 µL. For droplet formation, proteins at the indicated concentrations were tested for phase separation in buffer containing 20 mM Tris-HCl pH 7.4, and 100 mM KCl. The samples were added into the 384 well non-binding microplates (greiner bio-one). The images were taken after various time points starting from 30 minutes to 18 hours.

6.3.5 Image Collection

For the in-vitro assays, images were taken using an IX71/IX81 inverted Spinning Disc Microscopes with an Andor Neo sCMOS/Andor Clara CCD camera and an UPlanSApo 60x oil-immersion objective (Olympus).
6.3.6 Dynamic Light Scattering Experiments (DLS)

DLS measurements were performed using the Zetasizer Nano ZSP Malvern instrument (measurement range of 0.4 nm to 10 µm). The Nano ZSP instrument incorporates noninvasive backscattering technology. This enables the measurement of time-dependent fluctuations of the intensity of scattered light as scatterers undergo Brownian motion. The analysis of these intensity fluctuations enables the determination of the diffusion coefficients of particles, which are converted into a size distribution using the Stokes-Einstein equation \[32\]. The sample solutions were illuminated by a 632.8 nm laser, and the intensity of light scattered at an angle of 173° was measured using a photodiode.

In DLS, the autocorrelation function of the scattered light is used to extract the size distribution of the dissolved particles. The first order electric field correlation function of laser light scattered by a monomodal or monodisperse population of macromolecules can be written as a single exponential of the form:

\[
G(\tau) = 1 + b \exp \left( -2D_t q^2 \tau \right)
\]

Here, \( b \) is a constant that is determined by the optics and geometry of the instrument, \( D_t \) is the translational diffusion coefficient of the particles, and \( \tau \) is the characteristic decay time. The scattering vector \( q \) is given by:

\[
|q| = \frac{4\pi n_0}{\lambda_0} \sin \left( \frac{\theta}{2} \right)
\]

(6.5)
Here, \( n_0 \) is the refractive index of the solvent, \( \lambda_0 \) is the wavelength, and \( \theta \) is the scattering angle. For populations composed of a single type of scatterer, the distribution function of decay rates can be derived from a simple fit of the experimental estimates of the logarithm of the correlation function in Equation (6.4) to a polynomial. These methods, which apply to monomodal distributions of sizes of scatterers can be used to extract the translational diffusion coefficient, from which one can estimate the hydrodynamic radius \( R_h \) of the scatterers. For this, one uses the Stokes-Einstein relation:

\[
D_t = \frac{k_B T}{6\pi \eta R_h}
\]  

(6.6)

Here, \( k_B \) is the Boltzmann constant \((1.381 \times 10^{-23} \text{J/K})\) and \( \eta \) is the absolute (or dynamic) viscosity of the solvent. In this work, we used the hydrodynamic diameter \( d_h \) (i.e., \( d_h = 2R_h \)) as preferred way to quantify particle sizes.

**DLS Measurements**

All solutions were filtered using 0.2 \( \mu \)m membranes (Millex\textsuperscript{G}S units) purchased from Millipore\textsuperscript{TM}. All experiments were conducted with following settings on the Malvern instrument: Material - protein; Dispersant - 20 mM Tris buffers with 100 mM KCl salts; Mark-Houwnik parameters; Temperature: 25\( ^\circ \text{C} \) with equilibration time - 120 seconds, Measurement angle: 173\( ^\circ \). Each spectrum represents the average of 12 scans, each 10 seconds in duration. For every measurement, we recorded the autocorrelation function, intensity, and number.

All proteins were freshly purified and used after chromatography purification with standard stock solution buffer. For a typical measurement, the final buffer composition consists of 20 mM Tris 7.4 and 100 mM KCl. The samples were prepared by adding freshly prepared
stock proteins followed by dilution buffer, and mixed thoroughly by pipetting 4 to 6 times. The samples were equilibrated for 2 mins at 25°C and the data were recorded in 2-minute intervals.

### 6.3.7 Nanoparticle Tracking Analysis (NTA)

Nanoparticle tracking analysis was performed using NS300 from Malvern instruments (measurement range of 20 nm to 1 µm). The system was accompanied with a NanoSight syringe pump to inject the samples for the experiments. NTA measurements utilize the properties of light scattering and Brownian motion to obtain the size distributions and concentrations of particles in liquid suspension. A laser beam (488 nm) was passed through the sample chamber, and the particles in suspension were visualized using a 20x magnification microscope. The video file of particles moving under Brownian motion was captured using a camera mounted on the microscope that operates at 30 frames per second. The software tracks particles individually and uses the Stokes-Einstein equation resolve particles based on their hydrodynamic diameters.

All proteins were freshly purified and used after chromatographic purification with standard stock solution buffer. All buffers were filtered through 0.22 µm polyvinylidene fluoride membrane filter (Merck, Germany). All proteins stock solutions were centrifuged at 20000 RCF for 5 mins at room temperature prior to measurements. For typical measurements, the final buffer consists of 20 mM Tris 7.4 and 100 mM KCl. The samples were prepared by adding freshly prepared and centrifuged stock proteins followed by dilution buffer, and mixed thoroughly by pipetting 4 to 6 times. The samples were equilibrated for 2 minutes at 25°C and the data were recorded 6 minutes after sample preparation and equilibration.
6.3.8 Transmission Electron Microscopy (TEM)

TEM micrographs were acquired using Morgagni TEM (ThermoFisher) operated at 80kV with a Morada camera (EMSIS). Stock FUS-SNAP was diluted to 2 µM at 20 mM Tris & 4 and 100 mM KCl. 3 µL of sample was loaded on a carbon-coated copper grid, incubated for 6 mins, and then gently drawn off using a filter paper wick. This was followed by addition of 2 µL of 1.5% PTA pH 7.5. After 30 seconds the solution was gently drawn off by filter paper wick.

6.3.9 Förster Resonance Energy Transfer (FRET)

For FRET experiments, the SNAP tagged FUS proteins were mixed with SNAP-Surface Alexa Fluor 546 or 488 (NEB) at a 1:1.5 ratio at RT for 2 hr. Free dye was removed using Zeba Spin Desalting Columns (Thermo Scientific, Lot # QH222764) equilibrated with the storage buffer. FUS-SNAP containing 10% dye labelled FUS-SNAP-488 or FUS-SNAP-546 was mixed separately with buffer to yield a final protein concentration of 0.5 µM in 20 mM Tris 7.4 with 20 mM KCl. Then the solutions were mixed in equal volume. A 96 well plate (microplate, PS, half area, µClear, Med. binding, Black, Greiner Bio-one) was loaded with 100 µl of 0.5 µM FUS-SNAP 488, 0.5 µM FUS-SNAP-546, and the mixture. The spectra were recorded from 530 nm to 600 nm (10 nm bandwidth) with the TECAN plate reader using an excitation wavelength of 460±10 nm. For control, we used only Alexa 488 and Alexa 546 in similar range of concentrations with same buffer conditions. Additionally, we used high salt conditions (200 mM KCl) to dissolve the clusters and recorded the spectra at same settings.
6.3.10 Anisotropy Measurements

Setup

The anisotropy measurements were conducted on a confocal fluorescence microscope (FV1000 Olympus, Hamburg, Germany) using a polarized pulsed diode-laser (LDH-D-C-485, Pico-Quant, Berlin, Germany) at 485 nm. Laser light was directed into a 60x water immersion objective (NA=1.2) by a dichroic beam splitter and focused into the sample close to the diffraction limit. The light emitted was collected by the same objective and separated into two polarizations (parallel and perpendicular) relative to the excitation beam. The fluorescence signal was further divided into two spectral ranges (BS 560, AHF, Tübingen, Germany). Bandpass filters for eGFP fluorescence (HC 525/39) were placed in front of the detectors. The signal from single photon sensitive detectors (PDM50-CTC, Micro Photon Devices, Bolzano, Italy and HPMC-100-40, Becker&Hickl, Berlin, Germany, respectively) was recorded photon-by-photon with picosecond accuracy (HydraHarp400, PicoQuant) and analyzed using custom software (LabVIEW based). The temperature during all titration steps was 21.5 ± 0.5°C.

The anisotropy of Nile red with FUS-SNAP as titrant was measured with the same confocal setup using a supercontinuum laser (SuperK Extreme, NKT Photonics, Birkerød, Denmark) at 514 nm. The spectral ranges were separated by dichroic beamsplitters (BS 560 and 630 DCXR, AHF, Tübingen, Germany). Bandpass filters for eGFP (HC 525/39) and Nile red fluorescence (HC 607/70, HC 715/120) were placed in front of the detectors.
Titration Procedures

FUS-SNAP stained with 100 nM FUS-eGFP (both prepared using method B) were measured in a 50 mM Tris-HCl (pH 7.6) solution on a FUS coated cover glass. To account for the increasing KCl concentration with FUS-SNAP titration Tris-HCl buffer was added accordingly keeping the salt content constant until the saturation concentration resulting in a total eGFP dilution of 38 v%. FUS-SNAP was titrated into 20 nM Nile red solution at 50 mM KCl and 50 mM Tris-HCl (pH 7.6 at 21°C) whilst keeping both dye and salt concentration constant with Nile red containing buffer. The brightness of Nile red increases in a hydrophobic environment due to longer fluorescence lifetime shifting the emission spectrum to shorter wavelengths. Due to the direct binding of Nile red to FUS the anisotropy reflects the actual rotational diffusion of FUS oligomers not obscured by dye-linker motions as seen in the eGFP measurements.

6.3.11 Microfluidic Confocal Spectroscopy (MCS)

Microfluidic devices, shown by the design in Figure B.7 [33], were first fabricated as SU-8 molds (MicroChem) through standard photolithographic processes, and then produced as polydimethylsiloxane (PDMS) slabs, which were bonded onto thin glass coverslips [Xia1998](4). The devices were operated by placing gel-loading tips filled with buffer and protein sample in their corresponding inlet ports (Figure B.7B) and pulling solution through the devices in withdraw-mode at a flow rate of 150 µL/h using automated syringe pumps (neMESYS, Cetoni).

All experiments were conducted with FUS-eGFP fusion protein prepared using method A. The protein, stored in 500 mM KCl, 20 mM TRIS-HCl pH 7.4, was diluted with buffers of 20 mM TRIS-HCl to the indicated protein and KCl concentrations as stated. During the
experiment, the sample was placed into the sample inlet of the device (Figure B.7A) and the corresponding buffer containing the same concentration of KCl and TRIS-HCl into the buffer inlet. The co-flowing buffer was supplemented with 0.05% Tween-20 to prevent surface sticking of the protein to PDMS and glass surfaces.

Experiments were conducted by scanning the confocal spot of a custom-built confocal microscope through the central four channels of the microfluidic device (insert in Figure B.7A). A schematic of the optical unit is shown in Figure B.7B. Briefly, the setup is equipped with a 488-nm laser line (Cobolt 06-MLD) for excitation of GFP fluorophores and a single-photon counting avalanche photo diode (SPCM-14, PerkinElmer) for subsequent detection of emitted fluorescence photons. Further details of the optical unit have been described previously. During the scanning of the device, 200 evenly spaced locations within the central four channels of the device were surveyed and detected for 4 seconds. Examples of individual 4 second traces both with and without FUS clusters are shown in Figure B.7C. Clusters were classified as peaks that exceeded 5 standard deviations above the mean fluorescence intensity of each trace. These peaks were quantified according to location (Figure B.7D; top panel) against the mean signal of each trace (Figure B.7D; bottom panel). The average number of clusters was then quantified by averaging each of the four groups of peaks, corresponding to the four central channels. This was used in the calculation of cluster concentration according to:

\[ F_{\text{total}} = \left( \frac{\bar{n}_{\text{clusters}}}{t} \right) \times \left( \frac{h \, d_{\text{step}}}{\frac{\pi}{4} \, z \, w} \right) \]  

(6.7)

Here, \( t \) is the time each trace was collected for (4 seconds), \( h \) is the height of the microfluidic channel (28 µm), \( d_{\text{step}} \) is the width of each step (5.64 µm), and \( z \) and \( w \) were the height and width of the confocal spot (3 µm and 0.4 µm, respectively). From Equation (6.7), which yields the flux of clusters \( F_{\text{total}} \), the concentration of clusters could be determined according
to Equation (6.8), with $Q_{\text{sample}}$ being the flow rate of the sample (15 $\mu$L/h) and $N_A$ being the Avogadro constant [34](5):

$$c_{\text{cluster}} = \left( \frac{F_{\text{total}}}{N_A Q_{\text{sample}}} \right)$$

(6.8)

### 6.3.12 Computational Modeling

Simulations were performed using a customized version of the LaSSI simulation engine [28], and Chapters 2 and 3. For each of the simulations, a total of $N_{\text{tot}} = 2500$ polymers were used. Twenty different concentrations were sampled and these, written in terms of volume fractions, range from $\phi_{\text{min}} = 2 \times 10^{-5}$ to $\phi_{\text{max}} = 5 \times 10^{-2}$. The number of sites, $L$, on the $L \times L \times L$ cubic lattice was decreased from 1013 to 84 to increase the concentration by over three orders of magnitude. Radial density profiles from the centers-of-mass of the largest droplets were used to calculate the coexisting concentrations $\phi_{\text{sat}}$ and $\phi_{\text{den}}$, similar to Sub-section 4.3.3. Cluster size distributions were calculated from the simulations by quantifying the cluster composition of the system where a criterion of a maximal distance of $\sqrt{3}$ is used.

The basic architecture of the two models is the same: polymers that contain 12 beads connected by tether-like linkers of length 2. For the Homopolymer, all beads are of the same type, while for the Associative Polymer we have both stickers and spacers, as shown in Figure 6.14.

<table>
<thead>
<tr>
<th>Move</th>
<th>Frequency (%)</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anisotropic Small Cluster</td>
<td>0.24</td>
<td>10</td>
</tr>
<tr>
<td>Anisotropic Large Cluster</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>Proximity Small Cluster</td>
<td>0.24</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 6.2: Frequencies of different Monte Carlo moves. Note that in the second column the frequencies are normalized with respect to the lowest frequency – Proximity Large Cluster move in this case.

Simulation Protocol

All simulations start with random initial conditions. For $t_{EQ} = 5 \times 10^7$ MC steps, the simulation temperature is $T_{EQ} = 5T^*$. A constraining potential is applied to the system:

$$V(\vec{r}, \Delta T) = \Delta T \cdot (\vec{r} - \vec{r}_{cen})^2, \quad (6.9)$$

where $\vec{r}_{cen} = (L/2, L/2, L/2)$ is the center of the box, $\Delta T = T_{EQ} - T_1$. $T_1 = 1T^*$ corresponds to the first target temperature. This potential condenses all the chains to the center of the lattice. For Model B, the anisotropic interactions are turned off during this time.

After $t_{EQ}$ MC steps, anisotropic interactions are turned on, and the simulation temperature is exponentially annealed as such $T(t) = T_1 + T_{EQ}e^{-4t/t_{EQ}}$. Since $V(\vec{r}, \Delta T)$ depends on $\Delta T$,
the annealing also results in the relaxation of the constraining potential. The simulations are run for $t_1 = 1.25 \times 10^{10}$ MC steps (first cycle) after which the simulation temperature is discontinuously increased to $T_2 = 2T^*$ and run for another $t_2 = 1.25 \times 10^{10}$ MC steps (second cycle).

Data are acquired in the last half of each cycle at a frequency of $f_{\text{data}} = 2.5 \times 10^6$ MC steps, resulting in 2500 samples for each condition per run. The average of those 2500 samples is reported at the end of each simulation. The standard error of the mean between replicates is used as a measure of uncertainty. Lastly, data plots are generated using Python packages NumPy, Matplotlib and Seaborn. Adobe Illustrator is used to make the figures.

6.4 Results

6.4.1 Macroscopic Phase Separation Is Not Observed In Subsaturated Solutions

We quantified $c_{\text{sat}}$ using a spin-down based absorbance assay. At low concentrations, any prior aggregates that are present in solution are removed upon centrifugation at 20,000 relative centrifugal force (RCF) units. Results from this assay are summarized in Figure 6.1A for FUS-SNAP. At low $c_{\text{tot}}$, the concentration in the supernatant increases monotonically with the bulk concentration of SNAP-tagged FUS. However, above a threshold concentration, which we designate as $c_{\text{sat}}$, the concentration in the supernatant remains fixed at a plateau value. This is consistent with the establishment of phase equilibrium between dilute and dense phases (Figure 6.1A).
Figure 6.1: FUS-SNAP has a quantifiable $c_{sat}$ whereby condensates form only in supersaturated solutions. (A) Sample data for absorbance-based spin-down assays. Data shown here are for FUS-SNAP in 20 mM Tris, pH 7.4, and 100 mM KCl at $\approx 25^\circ$C. The red dashed line intersects the abscissa at 3 $\mu$M, which is the inferred $c_{sat}$ for this construct. Panels (B)-(F) show microscopy images collected at the 18-hr time point for solutions containing different concentrations of FUS-SNAP in 20 mM Tris, pH 7.4, and 100 mM KCl at $\approx 25^\circ$C. For imaging purposes, 5% of the total mixture in each sample is made up of FUS-eGFP. The scale bar in each panel corresponds to 10 $\mu$m. The data show the presence of condensates at or above $c_{sat}$ of $\approx 3$ $\mu$M. The FUS constructs used were expressed and purified using Method A (see Sub-section 6.3.2).

To test if $c_{sat}$ is a true saturation concentration, we performed microscopy-based measurements for solutions containing different amounts of FUS-SNAP and untagged FUS, Figure B.1). The concentrations investigated range from 0.25 $\mu$M to 4 $\mu$M. Data were collected approximately 30 minutes after sample preparation and for each sample, a series of images were collected at different time points over an 18-hour period. Results at the 18-hour time point are shown in Figure 6.1B-F for different concentrations of FUS-SNAP. Irrespective of the time point of interrogation, condensate formation of FUS-SNAP is only detectable at or above $c_{sat} \approx$
3 \mu M. Similar data were obtained for untagged FUS (Figure B.1). Based on these results, we conclude that macroscopic phase separation is not realized in subsaturated solutions and that the FUS molecules are defined by construct- and solution condition-specific saturation concentrations.

6.4.2 Clusters Spanning A Range Of Sizes Form In Subsaturated Solutions Of FUS

First, we used dynamic light scattering (DLS) to characterize subsaturated solutions of untagged FUS (Figure 6.2A, Figure B.2) and FUS-SNAP (Figure 6.2B-C, Figure B.3). Unless otherwise specified, all measurements were performed in 20 mM Tris, pH 7.4, and 100 mM KCl at \approx 25^\circ C. The DLS results we obtain are robust to the protocols used to express and purify FUS and FUS-SNAP molecules (Figure B.4).
Figure 6.2: DLS data show that clusters spanning a range of sizes form in subsaturated solutions. (A) Measurements were performed at different bulk concentrations of untagged FUS (see legend) that represent different degrees of subsaturation $-2 \leq s \leq 0$. Scattering intensities, measured 8 minutes after sample preparation, shift toward larger values as the bulk concentration $c_{\text{tot}}$ approaches the $c_{\text{sat}}$ of $\approx 2 \mu\text{M}$. Scattering intensity was measured at 0.5 $\mu\text{M}$ (B) and 1 $\mu\text{M}$ (C) of FUS-SNAP at 8-minutes after sample preparation and before centrifugation (brown curves). The solutions were centrifuged at 20,000 RCF for five minutes at room temperature. Then scattering intensities of the supernatant were measured at the 8-minute time point. These data are shown in blue curves in panels (B) and (C). They show the presence of species of sizes ranging between 7 nm and 100 nm. The FUS constructs were expressed and purified using Method A (see Sub-section 6.3.2).

The DLS data are shown as scattering intensities plotted against the apparent hydrodynamic diameter ($d_h$). For calibration, the $d_h$ of untagged FUS monomers is $\approx 7$ nm. The intensity profiles for untagged FUS collected at 0.25 $\mu\text{M}$ show two peaks (Figure 6.2A). The peak at $\approx 8$ nm corresponds to monomers and oligomers. The second peak at $\approx 200$ nm corresponds to mesoscale clusters. The observed bimodality could imply clusters of fixed size forming via either microphase separation [35] or micellization [36]. To test for this possibility, we investigated how the DLS signals change with increasing protein concentrations (Figure 6.2A). For microphase separation or micellization, the locations of the peaks should stay in place while the intensities of the peaks change respect to one another. We do not observe this behavior. Instead, the location of the peak corresponding to larger $d_h$ values shifts to the
right. At higher concentrations, DLS becomes blind to the presence of smaller species. As a result, the peak at lower values of $d_h$ vanishes above a concentration of 0.7 µM.

The apparent bimodality seen in intensity profiles can also arise from a heavy-tailed distribution of cluster sizes, whereby the most abundant species are monomers and oligomers. In this scenario, bimodality at low concentrations below 0.7 µM can result from smaller species being the most abundant, and the largest, mesoscale species having the highest scattering cross-sections [37]. This would mask the presence of species of intermediate sizes. To test for the presence of species of intermediate sizes, we subjected subsaturated solutions of FUS-SNAP molecules to centrifugation at 20,000 RCF. This removes species larger than 100 nm in diameter. Although, the bulk concentrations prior to centrifugation were 0.5 µM and 1 µM, respectively, the centrifugation step lowers the bulk concentration. The top of the centrifuged solution was then collected for DLS measurements. The DLS data, collected after centrifugation, are shown in panels B and C of Figure 6.2 for FUS-SNAP. The scattering intensity profiles, measured after centrifugation, show the presence of species in the size range of 7 – 100 nm for the apparent $d_h$. These species include monomers and an assortment of higher-order species. The upper limit on the number of FUS molecules per cluster ($n_{mol}$) is estimated as the ratio of the hydrodynamic volume of a cluster of size $d_h$ to that of the monomer. This yields upper bounds on estimates of $n_{mol}$ of $10, 10^2,$ and $10^3$ for $d_h$ values of $\approx 15$ nm, $\approx 30$ nm, and $\approx 70$ nm, respectively.

Taken together with data obtained prior to centrifugation, the parsimonious interpretation of the data is that the distributions of cluster sizes are likely to be exponentials with heavy tails such as Weibull distributions [38]. This would be concordant with reversible processes such as isodesmic associations that lack a threshold concentration with average sizes evolving continuously as concentrations increase [39, 40].

264
6.4.3 Morphologies Of The Largest Clusters That Form In Subsat-urated Solutions

We used negative stain transmission electron microscopy (TEM) to characterize morphologies of the largest clusters that form in subsaturated solutions of FUS-SNAP. Given their sizes, we refer to these species as mesoscale clusters. These clusters have roughly spherical morphologies (Figure 6.3A and Figure B.5). This was further confirmed using DLS data collected for untagged FUS at two different angles, Figure 6.3B. Based on their quasi-spherical morphologies and plausible densities within clusters, we estimate that mesoscale clusters with diameters of \( \approx 150 \text{ nm} \) encompass \( 10^2 - 10^3 \) molecules.

Figure 6.3: Mesoscale clusters that form in subsaturated solutions have quasi-spherical morphologies. Representative TEM image of 2 \( \mu \text{M} \) FUS-SNAP. The collections of TEM images were used to quantify the distribution of sizes (see inset and Figure B.6). (B) DLS data collected at two different angles corroborate the spherical morphologies of clusters observed in sub-saturated solutions. Samples used here were prepared using Method A (Sub-section 6.3.2).
6.4.4 Mesoscale Clusters Are Of Low Overall Abundance

We quantified the abundance of mesoscale clusters using nanoparticle tracking analysis (NTA) [41][37]. In these measurements, we track the Brownian motion of scatterers using dark field microscopy (Movies M1 and M2). We collected NTA data for untagged FUS (Figure 6.4A) and FUS-SNAP (Figure 6.4B). The distributions of mesoscale cluster sizes shift toward larger values with increasing protein concentrations. NTA also enables quantification of the percentage of molecules in solution that make up mesoscale clusters, Figure B.6). In a 250 nM solution of FUS-SNAP the bulk volume fraction of proteins is $\phi \approx 1.4 \times 10^{-5}$, and the fraction of the solution volume taken up by mesoscale clusters is $\phi_{\text{meso}} \approx 4 \times 10^{-8}$. This translates to 0.15% of the protein molecules in solution being a part of mesoscale clusters (Figure 6.4C). For a concentration just below $c_{\text{sat}}$, the relative abundance of mesoscale clusters increases to $\approx 1\%$. Similar results were obtained for untagged FUS, Figure B.6.

![Figure 6.4: Mesoscale clusters have low overall abundance.](image)

(A) NTA data for untagged FUS ($c_{\text{sat}} \approx 2 \mu\text{M}$) were collected at different concentrations that represent different degrees of subsaturation. (B) NTA data for FUS-SNAP ($c_{\text{sat}} \approx 3 \mu\text{M}$) at different degrees of subsaturation. (C) Relative abundance of mesoscale clusters, quantified at different degrees of subsaturation for FUS-SNAP. The constructs used were expressed and purified using Method A (Sub-section 6.3.2).
6.4.5 Clusters Increase In Size With Increasing Concentrations
And The Size Distributions Have Heavy Tails

To probe the distributions across a broader range of sizes, we used orthogonal, fluorescence-based approaches to characterize subsaturated solutions and access length scales that are distinct from those accessed using DLS and NTA measurements. Fluorescence anisotropy can be applied for detecting the presence of clusters with different numbers of molecules. These experiments use a small fraction of fluorescently labeled molecules (tracers) and a majority fraction of molecules that lack a fluorescent label. We used confocal multiparameter detection (MFD) which measures the fluorescence anisotropies of diffusing particles [42]. This allows us to monitor the size dependent rotational diffusion of molecules. We performed experiments with a fixed concentration (100 nM) of FUS-eGFP and added FUS-SNAP as a titrant with the concentration of FUS-SNAP ranging from 0 to 5 µM. The anisotropy increases monotonically at low concentrations of FUS-SNAP (Figure 6.5A). This indicates the presence of smaller clusters such as dimers and oligomers with increased volumes that result in longer times for molecular rotational diffusion. The anisotropy of eGFP is especially sensitive to the presence of small clusters. However, this signal saturates for larger clusters. This is because the measured apparent volume has a non-linear scaling due to the inverse relationship between anisotropy and volume. Further, the flexible FUS backbone and long eGFP linker may diminish the actual change in anisotropy. To test for the presence of larger clusters, we used the environmentally sensitive dye Nile Red, which offers a higher dynamic range in anisotropy. With Nile Red, we see a strong continuous increase in anisotropy as the concentration of FUS-SNAP is increased (Figure 6.5A). The signal does not saturate as the concentrations of FUS-SNAP increase. That this lack of saturation emanates from large bright clusters is readily confirmed in the binned histogram of fluorescence intensity of the photon trace (Figure 6.5B). These data show that the size distributions of clusters that form
in subsaturated solutions have heavy tails. The heavy-tailed distributions shift continually toward larger cluster sizes as concentrations increase.

Figure 6.5: Fluorescence anisotropy and brightness-based measurements show the presence of clusters in subsaturated solutions. (A) Normalized fluorescence anisotropy from confocal multiparameter detection (MFD) plotted against the concentration FUS-SNAP, which is the titrant. Data were collected at 50 mM KCl. The sample was prepared using method B. We also titrated FUS-SNAP from 0 to 3 µM in presence of 20 nM Nile red. Unlike the anisotropy measured using eGFP fluorescence (green), the measurements based on Nile Red (red) do not show saturation behavior, implying a continuous growth of cluster sizes with the concentration of FUS-SNAP. (B) With increasing FUS-SNAP concentrations in the presence of 20 nM Nile red, we observe bright particles, with apparent heavy-tailed distributions. (C) MCS data showing the formation of clusters in subsaturated solutions comprising FUS-GFP, prepared using method A. The apparent sizes of clusters, quantified in terms of photon intensity per cluster, increase with increasing protein concentration. The FUS-GFP constructs used in the MCS experiments were expressed and purified using Method A (see Sub-section 6.3.2).

Next, we turned to confocal detection under flow by using Microfluidic Confocal Spectroscopy (MCS). This is a brightness-based method that combines microfluidic mixing and flow with confocal detection [33, 34, 43]. MCS measurements are sensitive to the presence of species that span a broad size range from tens to hundreds of nanometers. Due to convective flow, it is not reliant on diffusion alone for sampling species of different sizes in solution (Figure B.7). We used FUS-eGFP ($c_{\text{sat}} \approx 4$ µM, Figure B.8) for the MCS measurements. Data were
obtained at concentrations of 0.25 μM, 0.5 μM, and 1 μM, plotted as the brightness per event, clearly show the formation of clusters in subsaturated solutions (Figure 6.5C).

6.4.6 Other FET Family Proteins Also Form Clusters In Subsaturated Solutions

Next, we asked if the formation of clusters in subsaturated solutions is unique to FUS or if this is a feature that is shared by other members of the FET family of RNA binding proteins. We collected DLS data in subsaturated solutions for untagged hnRNP-A3 ($c_{\text{sat}} \approx 6$ μM, Figure B.9A) (Figure 6.6A), EWSR1-SNAP ($c_{\text{sat}} \approx 2$ μM, Figure B.9B) (Figure 6.6B), and TAF15-SNAP ($c_{\text{sat}} \approx 2$ μM, Figure B.9C) (Figure 6.6C). These data show that other FET family proteins also form clusters in subsaturated solutions. The average sizes of these clusters increase with protein concentration. Raw DLS data in the form of autocorrelation functions are shown in Figures B.10–B.12.

Figure 6.6: Clusters also form in subsaturated solutions of FET family proteins. DLS data show evidence for mesoscale clusters in subsaturated solutions for (A) hnRNP-A3 ($c_{\text{sat}} \approx 6$ μM), (B) SNAP tagged TAF15 ($c_{\text{sat}} \approx 2$ μM), and (C) SNAP tagged EWSR1 ($c_{\text{sat}} \approx 2$ μM).
6.4.7 Clusters Are Reversible And Molecules Exchange Between Clusters

Using DLS, we find that the sizes of clusters decrease upon dilution and increase with increased concentration in the subsaturated regime (Figure 6.7A-C, Figure B.13). This observation points to the reversibility of cluster formation in subsaturated solutions whereby they shrink upon dilution and grow when concentrations increase. For cluster formation to be reversible, the molecules must exchange between clusters or between clusters and the bulk solution. To test for this, we used Förster resonance energy transfer (FRET) experiments (setup shown in Figure 6.7D). These experiments show that FUS molecules readily exchange across clusters (Figure 6.7E, Figure B.14).
Figure 6.7: Clusters in subsaturated solution form via reversible associations and molecules readily exchange between clusters. (A) DLS data were collected at different time points for SNAP-tagged FUS at 3 µM. (B) Upon dilution to 1 µM, the DLS data show changes to the intensity profiles, with the appearance of smaller species. (C) Increasing the concentration from 1 to 2.5 µM leads to an increased preference for larger species. (D) Design of the bulk FRET assay. Here, two sets of clusters are formed, each using a total concentration of 1 µM of SNAP-tagged FUS molecules. In each set, 5% of the molecules carry a fluorescent label (Alexa Fluor 488 - green or Alexa Fluor 46 - red). The clusters are mixed to achieve a total concentration of 1 µM. The mixture is excited using a 488 nm laser and the emission spectrum is measured from 520 nm to 600 nm. If molecules exchange between the clusters, then we expect to see a peak at the excitation maximum of 573 nm. (E) Fluorescence emission spectra show the decay of fluorescence for molecules with the Alexa 488 label and for the mixture. The latter shows a maximum at 573 nm, which is indicative of the exchange of molecules between the clusters. The FUS constructs used in this experiment were expressed and purified using method A.

The data presented in Figures 6.2–6.7 show the presence of clusters in subsaturated solutions. Cluster sizes increase as concentrations increase and approach $c_{\text{sat}}$. The clusters are equilibrium species that form and dissolve via reversible associations. The low abundance of mesoscale clusters quantified using NTA, the presence of smaller species, readily detected
using fluorescence anisotropy, and the presence of a broad spectrum of species indicated by MFD, fluorescence intensity analysis, and MCS suggest that the distributions of cluster sizes are concentration-dependent and heavy-tailed.

### 6.4.8 Growth Of Macroscopic Phases Above $c_{\text{sat}}$

In the physics of stochastic processes, it is known that random variables with broad distributions, specifically heavy-tailed distributions, are generators of condensation phenomena [44]. Therefore, guided by precedents in the literature [45, 46], we used DLS to ask if the heavy-tailed distributions of cluster sizes generate discernible signatures of transitions from finite-sized clusters to condensates just above $c_{\text{sat}}$. Specifically, we used DLS to probe the presence of slow modes [46] in the temporal evolution of autocorrelation functions for concentrations of untagged FUS that are below and above $c_{\text{sat}}$. Below $c_{\text{sat}}$, the autocorrelation functions reach a steady-state and do not change with time after a few minutes (Figure 6.8A). The timescales interrogated here are roughly seven orders of magnitude longer than the time it takes for individual FUS molecules to diffuse across 10 nm. Just above $c_{\text{sat}}$ (Figure 6.8B), the autocorrelation functions show the presence of slow modes. Such modes have been observed for polymers in dense phases, and have been attributed to reptation [46]. However, we interpret this to imply that clusters grow into micron-scale condensates above $c_{\text{sat}}$. The relevant data are shown for hnRNPA3, EWSR1-SNAP, and TAF15-SNAP (Figure 6.8C, Figure 6.9). That the slow modes in autocorrelation functions point to the onset of condensation processes was independently verified using microscopy. These data show that condensation as a function of time leads to the formation of dense phases (Figures 6.8 and 6.9).
Figure 6.8: Clusters grow into micron-scale bodies above $c_{\text{sat}}$. (A) Temporal evolution of autocorrelation functions from DLS measurements for untagged FUS. Below $c_{\text{sat}}$, the sizes of mesoscale clusters reach a steady state. (B) Above $c_{\text{sat}}$, we observe increased amplitudes of the autocorrelation function at longer times, (annotated using a gray arrow). (C) DLS data for hnRNPA3, EWSR1-SNAP, and TAF15-SNAP. Below $c_{\text{sat}}$ the cluster sizes reach a steady state. (D)-(G) Evidence of the growth into micron-scale condensates, displaying coarsening whereby fewer condensates grow by absorbing smaller species. This is made clear by the long-time evolution of micron-scale condensates formed by 3 $\mu$M untagged FUS containing 5% FUS-eGFP.

The observed temporal evolution is consistent with a coarsening process whereby larger condensates grow at the expense of smaller ones [47, 48]. Importantly, our analysis shows that the lowest concentration at which one observes the onset of slow modes in autocorrelation functions measured using DLS can be used as an efficient, centrifugation-free protocol for estimating $c_{\text{sat}}$. Further, these data suggest that $c_{\text{sat}}$ in systems that feature heavy-tailed distributions of clusters in subsaturated solution may be viewed as the threshold concentration for the onset of coarsening through coalescence, which is a signature of macroscopic phase separation [49].
Figure 6.9: Autocorrelation functions from DLS measurements collected at the protein-specific $c_{\text{sat}}$ values for (A) hnRNP-A3, (B) EWSR1-SNAP, (C) TAF15-SNAP, and (D) FUS-SNAP. These data show the presence of slow modes in the autocorrelation function profiles. These slow modes are only manifest at or above $c_{\text{sat}}$ and are absent below $c_{\text{sat}}$. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ$C.

### 6.4.9 Cluster Formation And Macroscopic Phase Separation Can Be Decoupled

Solutes such as 1,6-hexanediol (HD) can suppress phase separation and dissolve micron-scale condensates [31, 50]. We used DLS to measure the impact of increased HD concentration
on low abundance mesoscale clusters vs. macroscopic phase separation. We observe a dose-dependent response of macroscopic phase separation on HD. Specifically, the presence of a dense phase, characterized by the appearance of slow modes in the autocorrelation functions, is weakened and abrogated at increased concentrations of HD (Figure 6.10A-C). However, profiles of autocorrelation functions, namely the presence of fast modes that are consistent with the presence of clusters persist even upon the addition of up to 1% w/v of HD. Similar results were obtained when we queried the effects of ATP, which is thought to be a condensate dissolving hydrotrope at the concentrations used here [51]. Further, the distributions of clusters formed at $c_{\text{sat}} \approx 2 \mu\text{M}$ show minimal changes in the presence of HD and / or ATP (Figure 6.10D).
Figure 6.10: Solutes dissolve condensates while having a minimal effect on clusters. For untagged FUS, above $c_{sat}$, there are slow modes in the autocorrelation functions - see Figure 6.10B. This feature is weakened in (A) 0.25% and (B) 0.5% and lost in (C) 1% 1,6 hexanediol (HD), which dissolves condensates for concentrations above $c_{sat}$. However, clusters that form via fast modes persist upon the addition of HD. (C) Similar observations were obtained in the presence of 3 mM ATP-Mg. (D) NTA data, collected at a concentration of 2 $\mu$M for untagged FUS, show that the solutes have discernible effects on the distribution of cluster sizes. Within statistical error of the measurements, the effects of solutes on clusters in subsaturated solutions are small when compared to the impact of solutes on the loss of slow modes above $c_{sat}$.

We interpret the results of solute titrations to mean that there are at least two separable energy scales that contribute to condensate formation. Solutes such as HD and ATP primarily affect solvent quality, whereby the Flory $\chi$ parameter is altered to impact the overall solubility profiles of FUS molecules. This is consistent with the observations that polyols such as HD lower the macroscopic surface tension of water [52]. Accordingly, we reason that cluster formation is likely to be influenced by distinct, chemistry-specific interactions, that can be
separable using solutes but are generally strongly coupled to driving forces for macroscopic phase separation.

6.4.10 Impacts Of Mutagenesis On Cluster Formation And Phase Separation

We used a combination of mutagenesis experiments and experiments based on changes to solution conditions to query the extent of coupling between interactions that drive cluster formation vs. macroscopic phase separation. We measured the sensitivity of cluster formation and phase separation of full-length, untagged FUS to changes in pH (Figure 6.11). These experiments show that increasing the net charge weakens cluster formation in subsaturated solutions (Figure 6.11A, B). This leads to a clear upshift of the $c_{\text{sat}}$ for phase separation (Figure 6.11D). Likewise, increasing the concentration of monovalent salts also weakens cluster formation (Figure B.15). Therefore, increasing the net charge above a system-specific threshold or screening of electrostatic interactions weaken cluster formation and macroscopic phase separation.
Figure 6.11: Impact of pH and linked changes to net charge on cluster formation and macroscopic phase separation of untagged FUS. (A) Autocorrelation functions measured using DLS at the 8-minute time point for 1 µM untagged FUS. A series of traces are shown for solutions buffered to be at different pH values. The maximal amplitude and slowest decay are obtained for a pH of 9.0. (B) NTA data shown as distributions and as (C) abundance quantifying the distributions of mesoscale clusters at different pH values. (D) Estimates of the mean net charge of untagged FUS as a function of pH. These calculations made using unmodified pKa values for ionizable residues suggests that increasing the net charge weakens cluster formation. (E) Weakening cluster formation by increasing the mean net charge also weakens macroscopic phase separation. This is made visually evident using microscopy images, which show how condensates decrease/increase in size as the net charge increases/decreases. The scalebar is 10 µm and in the microscopy experiments 5% of the molecules are FUS-eGFP and the remaining 95% are untagged FUS. Samples were prepared using method A.
To test the contributions of chemistry-specific interactions to the coupling between cluster formation and macroscopic phase separation, we replaced 24 Arg residues in the RBD of full-length FUS with Gly residues. These mutations shift $c_{\text{sat}}$ up by at least an order of magnitude [10] and concomitantly lower the abundance of mesoscale clusters by over an order of magnitude (Figure 6.12A, Figure B.16A). Likewise, replacement of aromatic Tyr residues to Ser in the PLD of full-length FUS abrogates phase separation in the low micromolar range while also lowering the abundance of mesoscale clusters by over an order of magnitude (Figure 6.12B, Figure B.16B-D). These results highlight a strong coupling between clustering and macroscopic phase separation when Arg or aromatic residues are removed (Figure B.17, Figure B.18). In contrast, we find that substitution of 10 Asp and 4 Glu residues to Gly within the RBD of full-length FUS stabilizes cluster formation (Figure 6.12C, Figure B.16E). Specifically, at a concentration of 2 $\mu$M, the abundance of clusters increases by a factor of 1.8 when compared to wild type FUS. However, the increase in abundance of clusters is not accompanied by evidence of condensate formation even at a bulk concentration of 4 $\mu$M (Figure B.19).
Figure 6.12: Mutagenesis experiments reveal chemistry-specific effects of different residue types on the formation of mesoscale clusters in subsaturated solutions of FUS. (A) NTA data show that substituting 24 Arg residues to Gly within the RBD reduces the abundance of mesoscale clusters. (B) Replacing aromatic residues within the PLD of full-length FUS shows a valence-dependent reduction in the abundance of mesoscale clusters. (C) Replacing Asp / Glu with Gly in the RBD of full-length FUS, enhances chemistry-specific interactions that stabilize clusters. (D) Substitution of six Phe residues with Gly in the RBD of full-length FUS lowers the abundance of clusters. (E) Likewise, substitution of six Tyr with Ser in the RBD of full-length FUS also lowers the abundance of clusters.

Next, we queried the impact of replacing six of the Phe residues within the RBD of full-length FUS to Gly (Figure 6.12D, Figure B.16F) vs. replacing six of the Tyr residues in the RBD to Ser (Figure 6.12E, Figure B.16G). Although both Phe and Tyr are aromatic residues, the effects of replacing these moieties are quantitatively different. At equivalent concentrations, the abundance of mesoscale clusters, detected using NTA, is at least three-fold lower when Tyr residues are substituted with Ser when compared to substituting Phe residues with Gly. In both cases, macroscopic phase separation is also not observed for bulk concentrations up...
to 10 μM (Figure B.20). The differences between Phe and Tyr are reminiscent of recent results for the PLD of hnRNP-A1 [53].

6.4.11 Impact Of The Diversity Of Chemistry Specific Interactions

On Cluster Formation And Phase Separation

The distribution of residue types is different across full-length FUS compared to the PLD and RBD alone (Figure 6.13A). The measured $c_{sat}$ values of PLDs are at least two orders of magnitude higher than those of full-length proteins [10, 54]. We asked whether this change in $c_{sat}$ was accompanied by changes in cluster formation. NTA measurements show that at equivalent molar concentrations, the abundance of mesoscale clusters formed by the PLD of FUS is lower by a factor of 60 when compared to the abundance of such clusters formed by full-length FUS (Figure 6.13B, Figure B.21A). The DLS measurements show that macroscopic phase separation of the PLD was not observed for concentrations up to 150 μM (Figure B.21B). The sequence of the RBD features a significant number of aromatic and anionic residues and its measured $c_{sat}$ is ≈20 μM (Figure B.21C). This is an order of magnitude lower than that of the PLD. NTA data show that at a molar concentration of 1 μM, mesoscale clusters formed by the RBD (Figure 6.13C, Figure B.21D) are at least ten times more abundant than clusters formed by the PLD.
6.4.12 Simulations Show That Systems Featuring At Least Two Energy Scales Behave Very Differently From Those With A Single Energy Scale

One way to introduce distinct energy scales into a polymeric system is via the framework of associative polymers [10, 24, 25, 27–30, 55–59]. These are macromolecules with attractive groups known as stickers that are interspersed by spacers [24, 26, 27, 57, 60]. Stickers can form non-covalent reversible crosslinks with one another [31, 53, 61, 62]. The strengths of inter-sticker interactions are governed by the functional groups that define stickers. Therefore, in a stickers-and-spacers framework, inter-sticker interactions are specific interactions whereas spacers define the intrinsic $\chi$ and contribute to the overall solubility profile of the polymer [31, 53, 61, 62]. In recent computational work, FUS molecules were modeled as di-block polymers [63]. Note that associative polymers represent a superset that encompass specific instantiations such as di-block systems.
We used the LaSSI simulation engine, developed for lattice models of polymers [28], to model systems featuring a single energy scale vs. models for associative polymers that feature at least two energy scales. Our phenomenological models are intended to uncover possibilities that come from competing energy scales. First, we considered a model homopolymer where all units are identical, (Column 1 in Figure 6.14). Here, all interactions are uniform and isotropic, and they are described by single energy scale, $\epsilon$. For the finite-sized homopolymer considered here, phase separation requires that $\epsilon$ be negative ($\chi$ must be positive), and the magnitude of $\epsilon$ must be greater than $0.1k_B T$. Results for $\epsilon$ values of $-0.15k_B T$ and $-0.2k_B T$ are shown in rows 1 and 2 of column 1 in Figure 6.14. Making $\epsilon$ stronger by factor of $4/3$, realized by lowering $\epsilon$ from $-0.15k_B T$ and $-0.2k_B T$, decreases $\phi_{\text{sat}}$ by three orders of magnitude (column 1, rows 1 and 2 in Figure 6.14). This is because phase separation is a highly cooperative process when it is governed by a single energy scale [64]. Above a threshold value, small changes to the energy scale $\epsilon$ will dramatically alter the overall solubility, and this is manifest as large changes to $c_{\text{sat}}$. The dense phases formed by homopolymers are essentially polymer melts with volume fractions of $\phi_{\text{den}} \approx 1$ (Figure B.22).
Figure 6.14: Results from LaSSI simulations highlight the distinctions between homopolymers governed by a single energy scale vs. associative polymers with two energy scales. Each polymer has twelve beads. For homopolymers (results shown in column 1), all beads are the same. For associative polymers, the red beads are stickers, and the blue beads are spacers. We quantify the probability $P(n)$ of observing clusters comprising $n$ molecules. In each panel, the dashed lines represent distributions computed from LaSSI simulations that were performed above the model-specific $c_{\text{sat}}$, and the solid lines represent distributions computed in subsaturated solutions. The colors of the dashed and solid lines represent the extent of supersaturation ($s > 0$ for $\phi > \phi_{\text{sat}}$) or subsaturation ($s < 0$ for $\phi < \phi_{\text{sat}}$), respectively. Results in column 1 are for the homopolymer described by a single energy scale $\epsilon$, whereas results in columns 2-4 are for associative polymers featuring two energy scales namely, the excluded volumes of spacers and the specific sticker-sticker interactions with anisotropic interactions worth $g\epsilon$, where $g = 30$, 45, and 60, respectively.

The introduction of distinct energy scales alters the overall phase behavior. In our model for associative polymers, each bead is designated as being a sticker or a spacer (red vs. blue beads in Figure 6.14). Stickers form specific anisotropic interactions with one another. For a given value of $\epsilon$, the strengths of anisotropic inter-sticker interactions are set to be
where \( g \) equals 30, 45, or 60 (columns 2-4 in Figure 6.14). The spacers do not engage in attractive spacer-spacer or spacer-sticker interactions. Their only contribution is to the generic excluded volume. For values of \( \epsilon \) where the homopolymers undergo phase separation, we observe three distinct effects of introducing two different energy scales. First, the value of the saturation concentration \( \phi_{\text{sat}} \) is renormalized to be larger than that of the equivalent homopolymer. Although \( \phi_{\text{sat}} \) decreases with increasing \( g \), the decrease is considerably smaller when compared to what is observed for the homopolymer when the magnitude of \( \epsilon \) is increased by similar extents. Second, clusters form in subsaturated solutions. The size distributions of clusters are heavy-tailed, and their quantitative features are sensitive to both \( g \) and \( \epsilon \). This highlights the coupling between the effects of specific sticker-sticker interactions and generic, solubility determining interactions. Third, the interplay between stickers and spacers causes a dilution of polymer concentrations in the dense phase (Figure B.22). Instead of a polymer melt, the dense phases are condensate-spanning networks defined by inter-sticker crosslinks that are diluted by spacers. The overall thrust of our simulation results is concordant with results from simulations performed by Ranganathan and Shakhnovich of systems with similar architectures [65].

### 6.5 Discussion

A simple way to describe the driving forces for phase separation is using the mean field Flory-Huggins formalism [11, 12]. These simple theories have been the mainstay for quantitative descriptions and analysis of phase transitions driven by multivalent proteins and nucleic acids in vitro and in cells [13–16, 47, 49, 64, 66–70]. In this formalism, there is only one effective energy scale as captured by the parameter \( \chi \) [11]. For a purely phase separation-based process [11, 12], subsaturated solutions should feature dispersed monomers and very few small clusters, if any, at any given time. In this work, we report results from our investigations of the types
of species that form in subsaturated solutions of phase separating RNA binding proteins with disordered PLDs and RBDs. We find that subsaturated solutions do not conform to expectations based on systems that are characterized by a single energy scale such as the Flory $\chi$-parameter. Therefore, the biophysical principles underlying biomolecular condensate formation and dissolution should not be described using simple models for macroscopic phase separation.

The presence of different energy scales is best illustrated by the response to solutes, in which clusters and condensates respond differently to HD and ATP. Phase separation is likely driven mainly by composition-specific interactions that determine the solubility profiles of FUS and FUS-like molecules. However, our data also show that the sequence features of FUS and FUS-like molecules can engender a strong coupling between the driving forces for cluster formation and phase separation. Mutations to stickers weaken cluster formation and increase $c_{\text{sat}}$. This indicates that the mechanism of phase separation is intimately tied to the structure of the underlying size distribution of clusters that form in subsaturated solutions. The picture that emerges is of networks of chemistry-specific, inter-residue interactions [8, 10, 53, 54, 71] driving cluster formation and determining the extent of coupling between cluster formation and phase separation. It is important to emphasize that the clusters that we report here are not the same as pre-nucleation clusters that have been reported for various systems in supersaturated solutions [20–23]. Nor are they micellar species or facsimiles of microphase separation. Instead, they are distinct species that appear to form via mechanisms that are paralogous to isodesmic reversible associations [40], generating broad, heavy-tailed distributions of cluster sizes.

Recently, Zhao et al., showed that the N-protein of the SARS-Cov-2 virus, which has been studied for its macroscopic phase separation [72, 73], also forms an assortment of clusters in subsaturated solutions [74]. Seim et al., [75] discovered that the interplay between
specific homotypic and heterotypic interactions of the fungal protein Whi3 influences the distribution of protein cluster sizes that forms in dilute phases that coexist with Whi3-RNA granules. Mutations to the specific interactions that generate clusters through homotypic interactions can either enhance or diminish clustering, and this has a concomitant effect on the concentrations of Whi3 proteins in the dense phase [75]. Both reports highlight the importance of dilute phase clusters that are either modulated by heterotypic interactions or influence phase behavior driven by heterotypic interactions.

Mesoscale clusters of low overall abundance have been reported for subsaturated solutions of folded proteins that form crystalline solids [76, 77]. These clusters were found to have fixed size, irrespective of protein concentration [76], and are likely to be manifestations of microphase separation [35]. In contrast, we observe continuous growth in the sizes and abundance of clusters with concentration. As discussed below, the framework of associative polymers provides a plausible explanation for our observations.

Associative polymers undergo two types of transitions namely, percolation without phase separation - also known as sol-gel transitions - or phase separation coupled to percolation [26–29, 58, 59]. Above a system-specific threshold concentration known as the percolation threshold $c_{\text{perc}}$ [28–30, 58], non-stoichiometric clusters become connected into a percolated or system-spanning network [28–30, 58]. Quantitatively, $c_{\text{perc}}$ is determined by the numbers (valence) of stickers, diversity of sticker types [30], and the strengths of inter-sticker interactions (28, 29). If $c_{\text{perc}} < c_{\text{sat}}$, then associative polymers undergo percolation without phase separation. Alternatively, if $c_{\text{sat}} < c_{\text{perc}} < c_{\text{den}}$, where $c_{\text{den}}$ is the concentration of the dense phase formed via phase separation [59, 78], then phase separation and percolation are coupled (Figure 6.15). In this scenario, the condensates that form are akin to spherical microgels [79] whereby the percolated network is condensate spanning. The pre-percolation clusters that form in subsaturated solutions are direct manifestations of clusters predicted to define sols formed
by associative polymers undergoing either percolation without phase separation or phase separation coupled to percolation.

Figure 6.15: Cartoon Depicting The Concentration Dependence Of Systems That Undergo Phase Separation Coupled To Percolation. (A) When $c_{\text{perc}} < c_{\text{sat}}$ associative polymers undergo percolation without phase separation. As concentration is increased we get larger clusters until we reach $c_{\text{perc}}$, beyond which a system spanning network. (B) When $c_{\text{sat}} < c_{\text{perc}} < c_{\text{den}}$ associative polymers undergo both percolation and phase separation. As the total concentration of the system is increased, larger clusters are formed until we reach $c_{\text{sat}}$, after which the system phase separates. Since $c_{\text{den}} > c_{\text{perc}}$ the system forms networked droplets.

6.5.1 Precedents For Clusters In Subsaturated Solutions

The multivalent domain-linker systems studied by Rosen and coworkers are exemplars of associative polymers featuring an interplay between site-specific domain-motif interactions and linker-mediated solubility profiles [59, 78, 80–83]. Li et al. [78] investigated the phase behaviors
of systems where phase transitions are driven by multivalent interactions among poly-SH3 and poly-PRM molecules. Existence of a threshold concentration for phase transitions requires that there be at least three SH3 domains and three proline-rich modules (PRMs) within the associating molecules. This is consistent with the presence of a valence-dependent percolation threshold \[84, 85\]. Li et al. [78] showed that as the valence (numbers) of SH3 domains and PRMs increases, the threshold concentration for phase transitions decreases. Importantly, Li et al. [78] also reported the presence of clusters in subsaturated solutions. These were detectable using DLS and small-angle x-ray scattering. The phase transitions were referred to by Li et al., as macroscopic liquid-liquid phase separation that is thermodynamically coupled to sol-gel transitions. This phrasing is synonymous with the process we describe here as phase separation coupled to percolation.

### 6.5.2 Biological Relevance Of Clusters That Form In Subsaturated Solutions

Our results lead us to propose that RNA binding proteins from the FET family belong to the class of polymers known as associative polymers. Our findings show that chemistry- or site-specific sticker-mediated interactions and solubility determining spacer-mediated interactions are separable in systems with sticker and spacer architectures [29, 31, 53, 54, 58, 59, 86–92]. Pre-percolation clusters, with size distributions featuring heavy tails, are likely to be present at endogenous concentrations that tend to be in the nanomolar to low micromolar range in live cells [93]. The clusters we observe are likely to be the bridge that connects to recent findings regarding dynamic clusters comprising multivalent molecules [94, 95]. Importantly, the recent work of Cho et al., identified cytoplasmic clusters formed by several RNA-binding proteins under endogeneous, unstressed conditions [96]. These clusters are distinct from cytoplasmic stress granules, which are bona fide condensates that form in response to stress.
through the process of phase separation coupled to percolation [6–8]. Now that we know what to look for, we expect there to be many more accounts of clusters, precursors of condensates, forming at endogenous levels in live cells. The clusters of interest will lack a phase boundary, will form via reversible associations, and the distributions of their sizes and shapes will be determined by a complex network of homotypic and heterotypic interactions.

What are the potential consequences of there being clusters in subsaturated solutions in cells? Process control via biomolecular condensates requires that phase separation be robust and reproducible [4]. In a classical phase separation system, the barrier to nucleation will determine the response time. Clusters will likely speed up the response time by lowering the barrier for phase separation. Accordingly, we postulate that clusters poise RNA binding proteins for robust condensate formation in response to stimuli. Our discovery that heavy-tailed distributions of clusters exist in subsaturated solutions suggests that regulation can also occur at the level of shaping the distribution of cluster sizes. For example, chaperones are known to modulate size distributions of self-associating molecules [13], and one could imagine chaperones acting to reshape the size distributions of clusters that form in subsaturated solution and in “dilute” phases that coexist with condensates. Additionally, deleterious interactions with cellular components that lead to disease [97], and the dynamical arrest of condensate growth [65] could also occur at the level of clusters.
6.6 References


Chapter 7

Conclusion & Future Directions

7.1 Preamble

The research presented in this thesis helps in understanding the physical principles behind biology using phase separation as a possible means of spatio-temporal regulation of macromolecules. To generate insights regarding this phenomenon we developed a general and flexible MC simulation engine, LaSSI [1], which generalizes BFMs, and allows for flexible coarse-graining of biological molecules onto stickers-and-spacers architectures on a lattice (Chapter 2). After demonstrating that the methodology is physically sound, we applied it to a stylized instantiation of the protein FUS. We showed that we can generate full phase diagrams including percolation boundaries (Chapter 3). We consequently also learned that the concept of saturation concentration cannot be trivially extended to systems with multiple components – the dilute phase is described by a curve and not a point. Going further, we investigated how ligands offer one particular way to modulate phase behavior, and related the findings with Polyphasic Linkage as formalized by Wyman & Gill, [2] (Chapter 4). We found rich diversity
in how ligands affect the phase behavior of scaffold molecules, and that simply measuring
PCs for ligands was not enough to determine whether a ligand promoted or suppressed phase
separation. Moving on to investigate the possible functional implications of biomolecular
condensates we looked at the mt-Nucleoid (Chapter 5). The mt-Nucleoid was shown to be a
model condensate capable of actively transcribing RNA, and the complex multicomponent
phase behavior suggested that actively transcribing RNA causes the formation of novel
dynamically arrested morphologies. Finally, we refocused on the FET-family of proteins,
but in subsaturated solutions (Chapter 6). Counter to predictions by CNT, we found that
these subsaturated solutions were accommodate heterogeneous distributions of molecular
assemblies, the sizes of which could even reach the mesoscale. We further showed that the
driving forces for forming subsaturated clusters and phase separation were coupled, and that
there were multiple avenues to tune this coupling.

This concluding chapter is divided into two parts. The first focuses on summarizing Chap-
ters 2–6. The second part looks ahead, at possible new avenues of exploring the intricate and
exciting phase behavior of biomolecules.

7.2 Summary

Many biomolecular condensates form via phase transitions that are driven by multivalent
macromolecules. These molecules can be viewed as biological instantiations of associative
polymers that conform to a so-called stickers-and-spacers architecture. Stickers can be protein-
protein or protein-RNA interaction motifs or domains that form reversible, non-covalent
crosslinks with one another. Spacers are interspersed between stickers and their preferential
interactions with solvent molecules determine the cooperativity of phase transitions. In
Chapter 2, we discussed the development of an open source computational engine known as
LaSSI, [1], that can be used to calculate full phase diagrams for multicomponent systems comprising of coarse-grained representations of multivalent proteins.

We showed that LaSSI is designed to enable computationally efficient phenomenological modeling of spontaneous phase transitions of multicomponent mixtures comprising of multivalent proteins and RNA molecules. We then proceeded to demonstrate the application of LaSSI on linear, and branched, multivalent proteins in Chapter 3. We showed that dense phases are best described as droplet-spanning networks that are characterized by reversible physical crosslinks among multivalent proteins. We connected observations regarding correlations between apparent stoichiometry and dwell times of condensates to being proxies for the internal structural organization, specifically the convolution of internal density and extent of networking, within condensates. Finally, we wrapped Chapter 3 by demonstrating that the concept of saturation concentration thresholds does not apply to multicomponent systems where obligate heterotypic interactions drive phase transitions. Instead, the concentrations of all molecules determines the phase boundaries, and thus the dilute phase concentrations. This insight emerged from the ellipsoidal structures of phase diagrams for multicomponent systems and it has direct implications for the regulation of biomolecular condensates in-vivo.

With LaSSI in hand, we shifted focus towards investigations into possible ways to modulate the phase behavior of macromolecules. In Chapter 4 we took a deeper look at scaffolds, and how ligands offer one mechanism of modulation. Phase transitions of scaffolds can be regulated by changing the concentrations of ligands, which are defined as non-scaffold molecules that bind to specific sites on scaffolds. This mechanism of modulation, called polyphasic linkage, was formalized by Wyman & Gill, [2] and represents a framework for understanding how ligands can exert thermodynamic control over phase transitions. We used theory and computation to uncover some of the rules that underlie ligand-mediated control over scaffold phase behavior, [3]. Using a stickers-and-spacers model implemented in LaSSI,
we found that the modulatory effects of ligands are governed by the valence of ligands, whether they bind directly to stickers versus spacers, and the relative affinities of ligand-scaffold versus scaffold-scaffold interactions. Even more generally, we saw all ligands have a diluting effect on the concentration of scaffolds within condensates. We saw that while monovalent ligands always destabilize condensates formed via homotypic interactions, multivalent ligands can stabilize condensates by binding directly to spacers or destabilize condensates by binding directly to stickers. Furthermore, we showed that bipartite ligands that bind to both stickers and spacers can alter the structural organization of scaffold molecules within condensates even when they do not produce appreciable effects on scaffold coexisting concentrations. Lastly, we demonstrated that the $PC$ of the ligand alone convolves changes to coexisting concentrations and changes to phase volumes. We highlighted the importance of measuring dilute phase concentrations of scaffolds as a function of ligand concentration in cells. This could potentially reveal whether ligands modulate scaffold phase behavior by enabling or suppressing phase separation at endogenous levels, thereby regulating the formation and dissolution of condensates in-vivo.

In Chapter 5 we investigate the mt-Nucleoid, a model system for understanding condensates in the presence of active processes, and the production of a macromolecular species that then modulates the structures it was produced in. The functional consequences of concentrating cellular machinery into biomolecular condensates are still largely unclear. One fundamental cellular function that has been linked to condensate formation is transcription. Using in-vitro assays, we reconstituted mitochondrial transcription in condensates from purified components. We found that the core components of the mt-transcriptional machinery form multi-phasic, viscoelastic condensates in-vitro. Strikingly, the rates of condensate-mediated transcription are substantially lower than equivalent reactions in bulk solution. These condensate-mediated decreases in transcriptional rates are associated with the formation of dynamically arrested
vesicular structures that are driven by the production and accumulation of RNA during transcription. Using coarse-grained, equilibrium simulations, we showed that the generation of RNA alters the phase behavior and the organization of transcriptional components within condensates and that the in-vitro mt-condensates were more likely non-equilibrium structures. Together, our in-vitro and in-silico approaches helped us shed light on how proteins and (ribo)nucleic acids biophysically self-assemble within mitochondria in-vivo.

In Chapter 6 we investigated the subtleties of systems that undergo phase separation coupled to percolation. The dynamics of phase separation, especially at low endogenous concentrations found in cells, are thought to follow the tenets of CNT, describing a sharp transition between a dense phase and a dilute phase characterized by dispersed monomers. We used a multi-disciplinary approach combining theory, multiple experimental techniques, and computational modeling to understand subsaturated solutions of phase separating RNPs with intrinsically disordered PLDs and RBDs. Surprisingly, we found that subsaturated solutions were characterized by seemingly continuous heterogeneous distributions of clusters comprising tens to hundreds of molecules, and even a few mesoscale species that were several hundreds of nanometers in diameter. We found that these mesoscale clusters were fairly spherical and internally dynamic. The formation of clusters in subsaturated solutions is a shared property of FET family proteins. Furthermore, our results showed that cluster formation in subsaturated solutions, and phase separation in supersaturated solutions were strongly coupled via sequence-encoded interactions. We propose that the distributions of clusters formed in subsaturated solutions will be sequence-specific. Interestingly, however, we saw that cluster formation and phase separation could be decoupled from one another using solutes that impacted the solubilities of phase separating proteins, or they could also be decoupled by specific types of mutations. Inspired by the sequence differences between the RBD and PLD of FUS, we used LaSSI and demonstrated that multiple length and energy
scales enable clustering. We saw that for homopolymers with one only energy scale phase separation and percolation were highly coupled, but that this coupling could be modulated by introducing spacers and by having reversible pairwise interactions between stickers.

7.3 Future Directions

Subsaturated Solutions Of Associative Molecules

As we saw in Chapter 6, systems of biomolecules that can undergo density transitions coupled to networking transitions (phase separation coupled to percolation) are interesting even in the deeply subsaturated regime. While a lot of effort has gone into understanding the nucleation phenomenon of supersaturated solutions of biomolecules, [4–7], we hope that the subsaturated solutions receive the same systematic attention. Higher resolution\(^1\) data on the distribution of cluster sizes should begin to generate insights into the mechanistic details of clustering since this probability distribution of molecular species is the grand partition function.

Given our hypothesis that multiple energy or length scales are needed for enabling clustering, there are some basic questions worth asking. How many stickers would a polymer need to enable clustering? How many different types of stickers are required? Are simple 2-body like interactions sufficient to model such clustering? Towards tackling such questions, our current work is based on using LaSSI simulations to uncover the physical principles that underlie such subsaturated clustering. We wish to systematically understand how, given an associative polymer on the lattice (as in Figure 7.1), thermodynamically stable mesoscale assemblies can be generated.
Figure 7.1: A LaSSI model for understanding subsaturated clustering. (A) A single component polymer model where we further make the simplification that we have repeating blocks of 5 different monomers. This results in a $5 \times 5$ symmetric matrix where the diagonal represents “homotypic” interactions between monomers. (B) To determine how valency affects subsaturated clustering, we can restrict ourselves to just one monomer type, which also sets the interaction scale.

Initial results suggest that for 1-component systems, with 1 type of sticker, the total valency determines the clustering, as shown in Figure 7.2 row A, provided the inter-sticker interactions are sufficiently strong. With multiple sticker types, there needs to be a balance between the homotypic and heterotypic interactions, as shown in Figure 7.2 row B. Lastly, even beyond explicitly simulating such phenomenon, having purely theoretical insights into this behavior in a way that captures the sequence-encoded conformational ensemble at the nano-scale, the specific clustering distributions at the mesoscale, and the overall phase transitions at the micro-scale still remain a challenge.
Figure 7.2: Investigating subsaturated clustering with LaSSI. (A) With one interaction scale, $\epsilon = -10k_B T$, we see that the total valency determines the propensity to form larger clusters in the subsaturated regime. (B) When we have multiple interactions, a balance between the heterotypic and homotypic interactions is required to generate larger clusters in the subsaturated regime. As we go from the left to the right, we can think of the monomer interactions going from non-specific to very specific. As the interactions become more specific, clustering increases until we reach the rightmost case. Here, the interactions drive condensation instead.

Multicomponent Systems

As alluded to in Chapters 1–5, many biomolecular condensates encompass multiple components \textit{in-vivo}, [8–17]. While we know that the concept of a saturation concentration is not trivially transferable to multicomponent systems, [1, 9, 18], we still lack robust mechanistic insight into

\footnote{The high dynamic range of possible sizes, from $\sim 5$ nm to $\sim 500$ nm, certainly complicates such measurements using standard light scattering based techniques.}
how such systems behave. Even ligand-mediated regulation of the phase behavior of scaffolds represents a multicomponent system where the ligand is one of the components within the system. We have already started leveraging LaSSI to understand the physical principles that underlie such multicomponent systems. Figure 7.3 shows one particular instantiation of a multicomponent system. Here, the focus is on understanding how the phase behavior of a ternary system changes as we go from a system with purely heterotypic interactions to one where one component becomes an effective crowder.

Figure 7.3: A LaSSI model for understanding multicomponent phase behavior. Components A and B have the same polymer architecture. $\epsilon_{AA}$ is the strength of the homotypic interactions for A, while $\epsilon_{BB} = 0$ ensures that if the system phase separates we only get one droplet. $\epsilon_{AB}$ controls the heterotypic interactions between the two components. Here, $\delta \epsilon$ is a measure for the difference between the heterotypic and homotypic ineractions. As $\delta \epsilon$ goes from -1 to 1, we go from a system with only heterotypic interactions to one where only component A has interactions, and component B becomes a “crowder”.

As shown in Figure 7.4, we can see that the shapes of the phase diagrams are determined by the differences between the homotypic and heterotypic interactions. As we increase the homotypic interactions, the phase diagrams move from being well defined closed-loop curves to curves that intersect with the abscissa representing the component with homotypic interactions. Even more generally, we see that the slopes of the tie-lines are continuously rotated downwards.
These results are consistent with the theoretical predictions for multicomponent systems for associative polymers, as shown recently by Deviri et al., [18]. While these initial results seem to suggest that the difference between the homotypic and heterotypic interactions can drastically change the phase behavior, these data represent one particular polymer model, architecture, and interaction mechanism, and work is still to be done in understanding how the phase diagrams of multiple components can behave. Having experimentally measured complete phase diagrams, with tie-lines, for ternary systems would provide for a much needed test of such predictions.
Figure 7.4: Example phase diagrams for ternary systems comprised of linear multivalent associative polymers. Panels A – I: log-scale phase diagrams. As we go from panel (A) to panel (I), we go from a fully symmetric heterotypic only system to one where component B becomes an effective “crowder”. Therefore, in panel (I), the dense phase excludes component B.

1Measuring such phase diagrams is of course a problem in of itself. Even if we were to assume that somehow we have perfect measurements of the concentrations of every component in each phase, the combinatorial challenge of robustly sampling the now exponentially larger phase space is also a hurdle.
Lastly, it behooves us to reiterate that while understanding the equilibrium thermodynamic behavior of biomolecules is critical, biological processes are usually highly dynamic and far from equilibrium. The concentrations of molecules vary over time even for a single cell as part of the cell cycle, and furthermore, cells are not even thermally or chemically insulated from their environments. Differentiating between biomolecular condensates formed by true equilibrium processes and ones formed by non-equilibrium steady states will be important so that we can adopt the appropriate framework for further investigations. Even within equilibrium descriptions, the dynamic nature of biology begs us to also investigate the kinetics of such phase transitions. Polymeric systems can very easily get entangled, and not actually reach equilibrium within reasonable time scales. Indeed, not only has such dynamical arrest been touched upon in Chapter 5, but also the time dependent evolution of the material properties of condensates seems general, [19–24]. Similarly to how explicit molecular modeling allows for intuitive probing of changes in emergent properties due to changes in molecular interactions, field theoretic approaches allow for the inclusion of non-equilibrium effects and their effects on the emergent properties of the system. Owing to the ease with which one can think about and model non-equilibrium phenomenon, field theoretic approaches that can also tease out mechanistic details and the sequence specificity of the underlying molecules will prove to be very powerful.
7.4 References


Appendix A

Supplementary Tables For Chapter 4
Table A.1: Scaffold mutations associated with disease whose phase behavior changes from that of WT.
<table>
<thead>
<tr>
<th>Scaffold</th>
<th>Mutation</th>
<th>Disease</th>
<th>Effect</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HnRNPA1</td>
<td>D262V</td>
<td>Amyotrophic lateral sclerosis</td>
<td>Accelerates liquid-to-solid transition of hnRNPA1 droplet</td>
<td>Molliex <em>et al.</em> Cell 2015 [3]</td>
</tr>
<tr>
<td>Tau</td>
<td>Tau441: P301L, P301S, ΔK280, A152T</td>
<td>Alzheimer’s disease</td>
<td>Makes droplets under conditions in which WT tau441 does not</td>
<td>Wegmann <em>et al.</em> EMBO J 2018 [7]</td>
</tr>
<tr>
<td>Scaffold</td>
<td>Mutation</td>
<td>Disease</td>
<td>Effect</td>
<td>Citation</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>--------------------------------------</td>
<td>-------------------------------</td>
</tr>
</tbody>
</table>
Table A.2: Known ligands and their effect on scaffold phase behavior.
<table>
<thead>
<tr>
<th>Scaffold</th>
<th>Ligand</th>
<th>Effect</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kapβ2</td>
<td>Recovers WT FUS phase separation properties when disease related mutants promote phase separation (R244C) and aging (G156F)</td>
<td>Niazi et al. Mol Cell 2020 [12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delayed fibril formation and condensate hardening that is accelerated with the G156E mutation in vitro, Reduced motor defects in flies expressing P525L or R521C FUS</td>
<td>Wheeler et al. bioRxiv 2019 [13]</td>
</tr>
<tr>
<td></td>
<td>PR30</td>
<td>Promotes FUS prion-like domain phase separation in vitro</td>
<td>Boeynaems et al. bioRxiv 2020 [16]</td>
</tr>
<tr>
<td>Scaffold</td>
<td>Ligand</td>
<td>Effect</td>
<td>Citation</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>G3BP1 / RNA</td>
<td>BSA</td>
<td>Destabilizes phase separation and reduces concentration of hnRNP1 Δhexa in the dense phase with increasing BSA concentration</td>
<td>Kedersha et al. JCB 2016 [17], Sanders et al. Cell 2020 [20]</td>
</tr>
<tr>
<td>Huntingtin</td>
<td>mRNA</td>
<td>Promotes PGL-3 phase separation such that it phase separates at its physiological concentration</td>
<td>Sittler et al. Mol Cell 1998 [24], Saha et al. Cell 2016 [25], Ghosh et al. PNAS 2019 [26]</td>
</tr>
<tr>
<td>SH35/PRM5</td>
<td>Heparin</td>
<td>Promotes phase separation but then destabilizes phase separation at high heparin concentrations</td>
<td></td>
</tr>
<tr>
<td>Scaffold</td>
<td>Ligand</td>
<td>Effect</td>
<td>Citation</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>SH35/PRM5</td>
<td>Lysozyme</td>
<td>Destabilizes phase separation</td>
<td>Ghosh et al. PNAS 2019 [26]</td>
</tr>
<tr>
<td>UBQLN2</td>
<td>Ub</td>
<td>Destabilizes UBQLN2 liquid-liquid phase separation</td>
<td>Dao et al. Mol Cell 2018 [28]</td>
</tr>
</tbody>
</table>

Table A.3: Known ligands and their effect on scaffold phase behavior.
A.1 References


Appendix B

Supplementary Figures & Data For Chapter 6
Figure B.1: Microscopy-based verifications of the existence of $c_{\text{sat}}$ for FUS molecules. (A) The top row shows microscopy images collected 30 minutes after sample preparation for different bulk concentrations of FUS-SNAP. In these experiments, 5% of the molecules are FUS-eGFP. Micron-scale species are only visible for concentrations above $c_{\text{sat}}$. The bottom row, which shows images collected after 4-hours corroborates the veracity of $c_{\text{sat}}$ that was estimated using an absorbance assay as described in the main text. (B) Images collected for untagged FUS, 30 minutes after sample preparation. The images, collected at three different concentrations corroborate the veracity of the estimate for $c_{\text{sat}} \approx 2 \mu M$ for untagged FUS. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at $\approx 25^\circ$C.
Figure B.2: Raw DLS data for untagged FUS in subsaturated solutions. Each panel has two columns. The column on the left shows the autocorrelation functions collected at different time points after sample preparation and the start of the data collection. The amplitudes of autocorrelation functions are indicators of the sizes of scatterers in solution. The timescale for decay of the autocorrelation function is useful for assessing the apparent hydrodynamic diameters of scatterers. Autocorrelation functions can be converted to intensity profiles. These profiles, shown in the right column, are typically dominated by the largest species in solution. Each row corresponds to a distinct bulk concentration as marked on the left. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at \( \approx 25^\circ\text{C} \).
Figure B.3: Raw DLS data for FUS-SNAP \((c_{\text{sat}} \approx 3 \mu\text{M})\) in subsaturated solutions. The FUS-SNAP molecules used in these experiments were prepared using method A. As in Figure B.2, each panel has two columns. The column on the left shows the autocorrelation functions collected at different time points after sample preparation and the start of the data collection. The amplitudes of autocorrelation functions are indicators of the sizes of scatterers in solution. The timescale for decay of the autocorrelation function is useful for assessing the apparent hydrodynamic diameters of scatterers. The autocorrelation functions, converted to intensity profiles, are in the right column. Each row corresponds to a distinct bulk concentration of FUS-SNAP as marked on the left. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at \(\approx 25^\circ\text{C}\).
Figure B.4: Raw DLS data for FUS-SNAP in subsaturated solutions. The FUS-SNAP molecules used in these experiments were prepared using method B. Comparison of data shown here to the data shown in Figure B.3 indicate that the presence of DLS detectable mesoscale clusters, and their concentration dependence, are robust to the methods used to prepare FUS-SNAP samples. Method B includes a step wherein the samples are treated with benzonase, a generic nuclease. This should degrade the presence of any trace RNA contaminants that might be present in the sample. The data with and without this benzonase treatment step are equivalent to one another. This suggests that the properties of subsaturated solutions that we interrogate in our experiments are due to self-associations of FUS molecules. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at ≈ 25°C.

Figure B.5: Additional TEM images for FUS-SNAP. The images shown in panels A, B, and C show a roughly spherical morphology albeit with undulations that are visible on the length scales being interrogated. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at ≈ 25°C.
Figure B.6: Concentration dependence of the abundance of mesoscale clusters formed by FUS and FUS-SNAP molecules as measured using NTA. Panels (A) and (B) show data for untagged FUS and FUS-SNAP, respectively. The samples were prepared using method A. To calibrate expectations, we also performed NTA measurements for polyethylene glycol (PEG 100 kDa). PEG is a water-soluble polymer. As such, it should show negligible self-associations as a function of increased concentration. This is confirmed by comparing the distribution of detectable scatterers measured by NTA for PEG-100 kDa obtained at different polymer concentrations. Within error, these traces are only minutely different from that of the buffer alone. The low volume fraction of scatterers, quantified in panel (D) is consistent with the predominantly monomeric nature of PEG-100 kDa. All samples were prepared in 20 mM Tris, pH 7.4, 100 mM KCl at $\approx 25^\circ$C.
Figure B.7: Working principles of microfluidic confocal detection (MCD). (A) Schematic of microfluidic device used in MCD experiments. The insert and red line indicate the actual location within the channel which was scanned by the confocal spot. (B) Diagram of the working principle of the experiment and the optical setup. A 488 nm laser was used to excite the GFP-tagged fusion protein FUS. Emitted photons were counted on an avalanche photo diode, yielding raw traces as shown in the inserted plot. (C) By scanning through the microfluidic channel, the number of clusters could directly be counted as spikes in the photon trace. (top) When no clusters are present, a constant fluorescence signal is emitted from the solution of monomer. (bottom) Once conditions are adjusted to form clusters, the clusters are detected as clear spikes above the background, which are quantified by applying a cut off to count peaks the exceed 5 standard deviations above the mean fluorescent signal. (D) Clusters were quantified according to their position within the central four channels of the device (top panel, channels indicated by the dashed rectangles) and grouped according to the peaks in the mean intensity (bottom panel). The average of the total clusters in each of the channels was used for subsequent cluster concentration calculations.
Figure B.8: Visual confirmation that $c_{\text{sat}}$ for FUS-eGFP prepared using method A is $\approx 4 \ \mu\text{M}$. Microscopy images are shown here for different bulk concentrations of FUS-eGFP collected 30-minutes after sample preparation in 20 mM Tris, pH 7.4, 100 mM KCl at $\approx 25^\circ\text{C}$
Figure B.9: Saturation concentrations were measured for four different proteins in 20 mM Tris, pH 7.4, 100 mM KCl at ≈ 25°C. The measured $c_{\text{sat}}$ values were visually confirmed using microscopy for (A) hnRNP-A3 ($c_{\text{sat}} \approx 6 \, \mu\text{M}$), (B) EWSR1-SNAP ($c_{\text{sat}} \approx 2 \, \mu\text{M}$), and (C) TAF15-SNAP ($c_{\text{sat}} \approx 2 \, \mu\text{M}$).
Figure B.10: Raw DLS data for subsaturated solutions of hnRNPA3 \( (c_{\text{sat}} \approx 6 \mu\text{M}) \). The data are shown as autocorrelation functions collected at different time points for different bulk concentrations representing different degrees of subsaturation. The amplitudes of the autocorrelation function are governed by the scattering intensities and the decay times are governed by the sizes of scatterers. The data clearly show that the sizes of mesoscale clusters increase with increasing concentration. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at \( \approx 25^\circ\text{C} \).
Figure B.11: Raw DLS data for subsaturated solutions of EWSR1-SNAP ($c_{\text{sat}} \approx 2 \, \mu M$). The data are shown as autocorrelation functions collected at different time points for different bulk concentrations representing different degrees of subsaturation. The amplitudes of the autocorrelation function are governed by the scattering intensities and the decay times are governed by the sizes of scatterers. The data clearly show that the sizes of mesoscale clusters increase with increasing concentration. Samples were prepared using method A. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ C$. 
Figure B.12: Raw DLS data for subsaturated solutions of TAF15-SNAP ($c_{\text{sat}} \approx 2 \mu M$). The data are shown as autocorrelation functions collected at different time points for different bulk concentrations representing different degrees of subsaturation. The amplitudes of the autocorrelation function are governed by the scattering intensities and the decay times are governed by the sizes of scatterers. The data clearly show that the sizes of mesoscale clusters increase with increasing concentration. Samples were prepared using method A. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ C$. 
Figure B.13: Establishing for the reversibility of cluster formation. (A) Schematic summarizing the protocol used to dilute and concentrate FUS-SNAP samples (prepared using method A) starting from a reference bulk concentration of 3 µM. (B) Autocorrelation functions collected using DLS for a 3 µM solution of FUS-SNAP. This sets the reference amplitude maximum of the and decay time for the autocorrelation function. (C) Dilution by a factor of two causes a diminution in the maximum and shortens the decay time of the autocorrelation function. This indicates that dilution decreases the average sizes of clusters. (D) Concentrating the solution by a factor of 1.66 leads to autocorrelation functions with a larger amplitude and longer decay times. This is indicative of the average cluster sizes increasing as the concentrations increase.
Figure B.14: Control experiments performed using free dyes. These experiments indicate that the observations of FRET signals for FUS molecules arise because of the exchange of labeled FUS molecules and not because of the dyes alone.
(A) NTA - 1 μM FUS with 20 mM Tris pH 7.4

(B) MCS - 0.5 μM FUS eGFP

(C) 4 μM FUS with 20 mM Tris pH 7.4
Figure B.15: Increased concentration of KCl weakens cluster formation and macroscopic phase separation. (A) The distribution of cluster sizes, measured using NTA, shifts toward smaller values as the concentration of KCl is increased. Data are shown here for 1 µM untagged FUS prepared using method A. (B) The abundance of mesoscale clusters formed by 1 µM untagged FUS decreases with increased concentration of KCl. (C) Comparative MCS data show that fluorescence spikes present at 100 mM KCl for 0.5 µM FUS-eGFP are lost in the presence of 500 mM KCl. (D) Loss of mesoscale clusters with increased KCl concentration causes a loss of macroscopic phase separation. This is shown using microscopy data collected at a concentration of 4 µM of untagged FUS. For a fixed protein concentration, increasing the concentration of KCl leads to a loss of condensates.
Figure B.16: Impact of mutations within the PLD and RBD on the concentration dependence of size distributions of mesoscale clusters formed by untagged FUS. Data are shown from NTA measurements. (A) Substitution of 24 Arg residues in the RBD of full-length FUS with Gly weakens cluster formation in subsaturated solutions. (B) Substitution of 10 Tyr residues with Ser in the PLD of full-length FUS shifts the cluster distributions toward smaller sizes. (C) Substitution of 18 Tyr residues with Ser in the PLD of full-length FUS shifts the cluster distributions toward smaller sizes. (D) Substitution of 27 Tyr residues with Ser in the PLD of full-length FUS shifts the cluster distributions toward smaller sizes. Taken together, the data in panels (B)-(D) show a valence-dependent response, whereby decreasing the number of Tyr residues in the PLD of full-length FUS has a concomitant decrease in the sizes of clusters that are formed. The chemistry-specific nature of the driving forces for cluster formation in subsaturated solutions is revealed using differences in distributions of mesoscale cluster sizes formed upon replacing six Phe residues in the RBD of FUS with Gly - panel (E) - vs. replacing six Tyr residues in the RBD of FUS with Ser - panel (F). And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at \( \approx 25^\circ C \).

Figure B.17: Impact of substituting 24 Arg residues within the RBD of FUS to Gly. The data shown here are raw autocorrelation functions, gathered at different bulk concentrations of FUS (24R-G). These data demonstrate the reduced amplitude of the autocorrelation function (left), and rapid decay when compared to the wild-type FUS. The autocorrelation functions are converted to intensity profiles (right) to show that smaller species are formed when the Arg residues within the RBD are replaced by Gly. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at \( \approx 25^\circ C \).
Figure B.18: DLS data complementing the NTA distributions shown in Figure B.17. These data show the impact of valence-dependent substitutions of Tyr residues in the PLD to Ser. (A) FUS (10Y-S), (B) FUS (18Y-S) and (C) (27Y-S). These data demonstrate the amplitude of the autocorrelation function (left), and the autocorrelation functions are converted to intensity profiles (right). And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ$C.
Figure B.19: Impact of substituting 10 Asp / 4 Glu residues to Gly within the RBD of FUS on the DLS measured autocorrelation functions at different concentrations corresponding to subsaturated solutions. When compared to the data from Figure B.2, the data shown in panels (A) - (D) suggest that the replacements of 10 Asp / 4 Glu residues to Gly within the RBD of FUS has a minimal impact of the extent of cluster formation in subsaturated solutions. However, panel (E) shows that the mutations weaken macroscopic phase separation as evidenced by the lack of slow modes being present above the $c_{sat}$ values that were quantified for wild-type FUS. These data show that mutations can maintain cluster formation while weakening macroscopic phase separation. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ$C.
Figure B.20: DLS data showing the chemistry-specific effects of Phe to Gly (A) vs. Tyr to Ser (B) substitutions on the formation of mesoscale clusters in subsaturated solutions. These data demonstrate the amplitude of the autocorrelation function (left), and the autocorrelation functions are converted to intensity profiles (right). And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at \( \approx 25^\circ C \). And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at \( \approx 25^\circ C \).
Figure B.21: (A) NTA data collected at a series of concentrations for the PLD compared to the data collected for untagged, full-length FUS at 1 µM. These data show that the PLD alone is a weak driver of mesoscale clusters. (B) The NTA data are confirmed using DLS, as shown using raw autocorrelation functions collected for concentrations of the PLD that range from 10 µM to 150 µM. (C) The autocorrelation functions, when converted to intensity profiles, highlight the dominance of smaller species for the PLD. (D) The autocorrelation functions and accompanying intensity profiles - panel (E) - obtained for the RBD of FUS stand in contrast to those for the PLD. The autocorrelation functions show the onset of slow modes at an RBD concentration of 20 µM. This implies the onset of macroscopic phase separation above a $c_{sat}$ of 20 µM, which is an order of magnitude lower than the $c_{sat}$ for the PLD at similar solution conditions. (F) NTA data show the robust presence of mesoscale clusters forming in subsaturated solutions of the RBD. And data were collected in 20 mM Tris, pH 7.4, 100 mM KCl, at $\approx 25^\circ$C.
Figure B.22: The volume fractions of polymers in dense phases formed for homopolymers vs. associative polymers. Dense phases formed by homopolymers with a single energy scale are akin to polymer melts, $\phi_{\text{den}} \approx 1$. The dense phases formed by associative polymers are diluted compared to the homopolymers. This is because of the effects of spacers and the arrangement of stickers into crosslinked networks. Note that the energy scales for observing phase separation are renormalized by the stereospecific interactions among stickers.

### B.1 Amino Acid Sequences Of Proteins Used

**FUS-eGFP**

```
MASNDYTQATQSYGAYPTQPQGYSQQSSQPYGQYSYGSSYQSTDTSYGQSSYSSYQGQNTGYGTQSTPQGYG
STGGYGSSQSSQSSYQGQPPASPSTSGYSSYQSSYQGQPPQGQPSYQSGSQQPSYGQPSYQSQSYPQYNPP
QGYQQQNYNSNSGSSGSSSSNSSYYQDSMSSSGSSGSSSYYQDSMSSSGSSGSSGSSSYYQDSMSSSGSS
GGGGYNRSSGGSYGEEPRGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGR
GGGGGNRRSSGGSYDPRGRGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGG
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
DYNKQGIKTKHKTGQPMNLYTQTQGKGATVSFFDDPPSAKAAIDWFDGKEFSGPNKDKVSMFATRARDFNRE
GNRGRGGRGMGRRGGGGYGGGGGSSGGGGGGGGGQRRAGDWWKCPNPTCENMFSWRNECNQCKAPKDPGG
QPMGSGMNGGGNCDRGGRRGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGGRGG
QGSMSGSOSLSGLGKMLKFGKLFKSKLKGKFLKFLKFLKLKFLKFLKFLKFLKFLKFLKFLKFLKFLKFLK
GVQCSYRDPHMKQHHFFKASMPGEGYQERTFFKKDDGNYYKTRAEVKFEYCEVTLTVNRIELKGIDFEDGNILGHKLE
```
FUS-SNAP

YNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPEKRDHMVLLEFVTAAGITLGMDELYK

FUS

YNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPEKRDHMVLLEFVTAAGITLGMDELYK
FUS-PLD(1-211)

MASNDYTQQATQSYGAYPTQPQGQYSQQSSQPYGQQQSYSFGYESQSTDTSQYGQSSSYSSYGQSQNTGYGTQSTPQGYG
STGGYGSSQSSQSSYQQPSYPYGQQPAPSSTSTGYGSSSSQSYSEQQPSQSGYQQQSYQESQSYGQQSYQSYNPP
QGYGQQNYNSSSGSGGGSOGGNNYQGDQSSMSSGGGSGGYYQGNDQSGGGGSGGYYGQQ

FUS-RBD(212-526)

DRGGRGRRGGGSSGGSSGGGYSYNRSSGSGYEPGRGGGRGGGRRGGGGOSSRGRGGFNKFQGRPRDQGSPRHDSEQDSNDNNTI
FVQGLGENVIESVADYFKQIGIITKTNKGTGQPMPINLYTRGTGKLKGEATVSGFDPPSAKAAIDWDGDKEFSGNP
IKVSFATRRADFRNRRGGNGRRGRRGGPMGRQGGYGGGSGGGGRRGFPSSGGGGGGQQRAGDWKCNPCTENMNF
WRNENQCKAPKDPGPGGPGGGHSMMGNYGDRRGRRGGYGRGDRGGFRGRGGRDPGPGKMDSRG
EHRQDRRERP

FUS-(RBD)24R-G

MASNDYTQQATQSYGAYPTQPQGQYSQQSSQPYGQQQSYSFGYESQSTDTSQYGQSSSYSSYGQSQNTGYGTQSTPQGYG
STGGYGSSQSSQSSYQQPSYPYGQQPAPSSTSTGYGSSSSQSYSEQQPSQSGYQQQSYQESQSYGQQSYQSYNPP
QGYGQQNYNSSSGSGGGSOGGNNYQGDQSSMSSGGGSGGYYQGNDQSGGGGSGGYYGQQ
GGGGYNNRSSGSGYEPGGOGGGGGOGGGMGGSGGDDGGFNKFQGRPRDQGSPRHDSEQDSNDNNTIFVQGLGENVIESVA
DYFKQIGIITKTNKGTGQPMPINLYTRGTGKLKGEATVSGFDPPSAKAAIDWDGDKEFSGNPIKVSFATRRADFRNGG
GGNNGGGGGGGPMGGGGYGGGSGGGGSSGGSFFGSQGQQRAGDWKCNPCTENMNFWRNENQCKAPKDPG
PGGPGGGHSMMGNYGDRRGRRGGYGGGDDGGGDDGGGGGPGKMDSGGEHRQDRRERP
FUS-(RBD)10D/4E-G

MASNDYTQATQSYGAYPTQPQQGYSQQSSQPYGQQSYSGYSSQSTDTSGYGQQSSSYGQNTGYGTQSTPQGYG
STGGYGSQQSSQSYGQQSYPGYQQPAPSTSGYSSQSSSSYQGPSQGSYQPSYQGQQSSQSYQSNPP
QGYGQNNQNYNSSGGGGGGGGGGNYQDQSSMSSGGGSSGYQADRGGRGGRGGGGGGG
GGGGYNRSYGYEPGRGGGCRGGRGGMGSGDRFFGNKFGPRDQGSRHDSEQDNSDNNTIFVQLGENVTIESVA
DYFKQIGIIKTNKTQMPMINLYTDRETGLKGEATYSDKDPPSAKAIAIDWFYDGKEFSNPKVSFATRAGFRG
GGNRRGGRRGMPRGGRGGYGGGGGGGGGGGGFFSGGGGGGGGQRRAGWKPNTCPGNNFWRNCNQCKAPKPG
PPGGPGGSHGGNYYGGGGGRRGQGGYGRGGGGFRGRRGGGFGPGKMBRSRGHRQGRRGRPY

FUS-(PLD)10Y-S

MASNDYTQATQSYGASPTQPQQGYSQQSSQPSGQQSYGSSQSTDTSGYGQQSSSYGQNTGYGTQSTPQGS
STGGYGSQQSSQSYGQQSYPGYQQPAPSTSGYSSQSSSSYQGPSQGSYQPSYQGQQSSQSYQSNPP
QGYGQNNQNYNSSGGGGGGGGGGNYQDQSSMSSGGGSSGYQADRGGRGGRGGGGGGG
GGGGYNRSYGYEPGRGGGCRGGRGGMGSGDRFFGNKFGPRDQGSRHDSEQDNSDNNTIFVQLGENVTIESVA
DYFKQIGIIKTNKTQMPMINLYTDRETGLKGEATYSDKDPPSAKAIAIDWFYDGKEFSNPKVSFATRAGFRG
GGNRRGGRRGMPRGGRGGYGGGGGGGGGGGGFFSGGGGGGGGQRRAGWKPNTCPGNNFWRNCNQCKAPKPG
PPGGPGGSHGGNYYGGGGGRRGQGGYGRGGGGFRGRRGGGFGPGKMBRSRGHRQGRRGRPY

FUS-(PLD)18Y-S

MASNDSTQATQSYGASPTQPQQGSSQSSSQQSPGQQSYSQSSQSTDTSGYGQQSSSYGQNTGSQGSTPQGS
STGGYSSQSSQSYGQQSYPGYQQPAPSTSGYSSQSSSSQGPSQGSYQPSYQGQQSSQSYQSNPP
QGYGQNNQNSSSGGGGGGGGGGNGSQQDQSSMSSGGGSSGYQADRGGRGGRGGGGGGG
GGGGYNRSYGYEPGRGGGCRGGRGGMGSGDRFFGNKFGPRDQGSRHDSEQDNSDNNTIFVQLGENVTIESVA
DYFKQIGIIKTNKKTGQMINLYTDRETGLKGEATVFSDDPGSAKAADWFDGKEFSGNPIKVSFAHTRADNFRG
GGNRRGGRGGRGPMGRGGYGGGGGGGGGGGGGQQRAGDWKCPNPTCENMNFSWRECNQCKAPKPDG
PGGGPGGSSHMGGNYGDDDHRGGGRGGYDRGGYRGGRGGDRGGGGPGKMDSRGHEHRQDRRERPY

**FUS-(PLD)27Y-S**

MASNDSTQQATQSSGASPTQPQGSSQSSQPSGQSSSGSSQQSTDTSGSQGSSSSSSGQSSQNTGSQTGQSTPQGS
STGGGSSQQSSQSSQSSQSPSGQQPAPSSTSGSSQSSSSQSSSQPSGSSSSQPSQGSSSSQPSGQSSGSSQGSSQSSNSP
QGGQQSQQNSSSSGGGGGGGGGGGGNSQDQSSMSSSGGGGGGGQDQGSSQGSSQGSSQSSQGSSQGSSQGSSQGSSQSSN
GGGGGNRGGRRGGGRRGGYGGGGGSSRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGG
PGGGPGGSSHMGGGYGDDDHRGGGRGGYDRGGYRGGRGGDRGGGGPGKMDSRGHEHRQDRRERPY

**FUS 6F-G**

AMASNDYTQQTQSYMAYPTQPGGGSSQSSQPSGQSSSYGSSSQSTDTSGYGSSSYGSSQSNQTGSGYGTQSTPQGY
GSTGGYGGSSQSSQSSYQGPSYPGYQAPPSSTSGSYGSSQSSYSQPSQGSSQPSQPSYGGQSSYQQSYQSYQPQNSPS
PQGYQQNYNSSSGGGGGGGGGGGGGGGGNYQDQSSMSSSGGGGGGGGNYQDQSSMSSSGGGGGGGGNYQDQSSMSSSGGG
GGGGGNRGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGGRRGGG
PGGGPGGSSHMGGNYGDDDHRGGGRGGYDRGGYRGGRGGDRGGGGPGKMDSRGHEHRQDRRERPY

[362]
FUS 6Y-S
AMASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGY
GSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNP
PQGYGQQNQYNSSSGGGGGGGGGGNYGQDQSSMSSGGGSGGGYGNQDQSGGGGSGGYGQQASDRGGRGRGGSGGGG
GGGGGGSNRSSGGSEPRGRGGGRGGRGGMGGSDRGGFNKFGGPRDQGSRHDSEQDNSDNNTIFVQGLGENVTIESV
ADYFKQIGIIKTNKKTGQPMINLYTDRETGKLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNR
GGGNGRGGRGRGGPMGRGGSGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKAPKPD
GPGGGPGGSHMGGNSGDDRRGGRGGSDRGGSRGRGGDRGGFRGGRGGGDRGGFGPGKMDSRGEHRQDRRERPY

TAF15-SNAP
MSDSGSYGQSGGEQQSYSTYGNPGSQGYGQASQSYSGYGQTTDSSYGQNYSGYSSYGQSYSQSYGGYENQKQSSYS
QQPYNNQGQQQNMESSGSQGGRAPSYDQPDYGQQDSYDQQSGYDQHQGSYDEQSNYDQQHDSYSQNQQSYHSQREN
YSHHTQDDRRDVSRYGEDNRGYGGSQGGGRGRGGYDKDGRGPMTGSSGGDRGGFKNFGGHRDYGPRTDADSESDNS
DNNTIFVQGLGEGVSTDQVGEFFKQIGIIKTNKKTGKPMINLYTDKDTGKPKGEATVSFDDPPSAKAAIDWFDGKE
FHGNIIKVSFATRRPEFMRGGGSGGGRRGRGGYRGRGGFQGRGGDPKSGDWVCPNPSCGNMNFARRNSCNQCNEPR
PEDSRPSGGDFRGRGYGGERGYRGRGGRGGDRGGYGGDRSGGGYGGDRSSGGGYSGDRSGGGYGGDRSGGGYGGDR
GGGYGGDRGGGYGGDRGGGYGGDRGGYGGDRGGGYGGDRGGYGGDRGGYGGDRGGYGGDRGGYGGDRSRGGYGGDR
GGGSGYGGDRSGGYGGDRSGGGYGGDRGGGYGGDRGGYGGKMGGRNDYRNDQRNRPYGAPGSSSGRENLYFQGMDK
DCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEF
PVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISYSHLAALAGNPAATAAVKTALSGNPVPILIPCHRVVQGDLDV
GGYEGGLAVKEWLLAHEGHRLGKPGLG

[363]


EWSR1-SNAP

MASTDYSTYSQAAAQQGYSAYTAQPTQGYAQTQAYGQQSYGTYGQPTDVSYTQAQTATAYGQTAYATSGYQQPPTG
YTTPTAPQAYSQVPQVGYGTGAYDTTTATTTTTAQQSYADDQQSAYGTQPAYPAYGQQPAATAAPTRPQDGNKPTETSQPQ
SSTGYNYQPSLGYQSNYSYPQVPGSYPQMPQVTAPPSYPTSTYSTQPTSYDQQSSYSQNTYGPSSYQQSSYQQ
QSSYQQPPPTSYPPPTGSYSQAPSQYSQQSSSYQQSSSYQQSSSYQQSSSFQRDHPSSMGVYGPESGGFSGPGGENRSMG
PSGPDNRGGR
GGFDGRGMSRGGRRGGGMRGSMAGERGGFNGKPGPMDEPGLDLGPLVDPDEDSDNSAYVQLNDSDLDDLADF
FKQCQVVKMKNKTGQPMIHYLDKETGPKGDATVSYEDPTAKAAVEWFDGDQGSKLKVSLARKKPPPMSMGR
GLPPREGRGMPPLRPQPQPQPQGGDRGPGGGRGGRGDRGGRGFRGRGGGDRGGRGMDRGQ
GGGGRRGPGGPGPGPGLMEQMGGRGGGRRGGPGKMDKGEHRQERRDRPYGAPGSSGRENLQFGMDKDCMKTTL
DSPGLGKLELSGCEQQLHRIIFLGKGTAADAVEVPATAVLGGPPEMMDGQAATLNVAYFHQPEAISEFVPALHHPV
FQQESFTRQVLWLKLKVFKFGEVISYSHLAALAGNPAATAAVKTAALSGNPVIPILCHVRVQVDLGDGGEGELAV
KEWLLAHEGRHLKPGPLG

hnRNPA3

MEVKPPGRPQPDGSNGRRRRRGEUGHDPKDEPEQLRLGIGGLSFEOTTDSLREHFGEKWGLTDVVMDPQTKRGR
GGFVTVSCVEEVDAAACMPHVKVGRVPEKRAVRSREDVSKPGAHTVVKIVGIGKEDTEEEYNLVRDYEFEKYIK
ETIEVMEDRQSGKRGAFVTFDDHDTVDIVQKYHTINGHNCEVVKALKSKQEMSAGSQRGRGGGSNFMRRG
NFGGGGGFGRGGNFGGGGGRGGGGGSSSSGGGGSSGGGYGGGGGSPGSSGGGGGYGGGSPGSGSGGGGGY
GGGGYDGYNEGGFNGGGNNYGGGGNNDFQNGRNYSQQQSGNGMKGGSGSFGGRSSGPGPYYGGGGSFGGSY

[364]
TDP-43-eGFP

MSEYIRVTEDEDEEPDIEDGTVLLSTVTAQFPGACGLRNYRNPVQCMRGVRLLVEGILHAPDAGWNLYVYVYN
YPKDNKRKMDETDASSAVVKRAVQKTSVDLIVGLPWTQDLDLEYFSTFGEVLMQVKKDLKTHGSTKGFGFVRFP
TEYETQVAKMSQVHRIDGRCDCCLPLNQSQDEPLRSRKVFVRCEDMETEDELREFFSQYGDVMDFVIPKRPFA
FAFVTFADEQIIAQSCLCEDLIIGISVHSNAEPRKHSVRLGRSFRGGNPGGFGNMQGFGNSRGGGAGLGNQG
SNMGGGMNGFASINPAMMAAAQALQSSWGMMLASQQNQGNSQGNNQMNQGMNPQNPQAFAFGNSYSGNSNG
AAIGWGSANAGSGFTGSGSSMDKSSWGMAGPSGSSRENLYFQGMVSKGEELFTGVVIPVLDGVDVNGH
KFSVSGSEGEGDATYGKTLKFICTTGLPVPWPLVTLTLYGVCFSRYPDHMQHDRFKSAMPEGYVQERTIFFK
DDGNYKTRAEVKFDGTLVNIKLKIDFKEDGNIILGHKLEYNYNHVYIMADKQKNGIKVNFKIRHNIEDGSVQ
LADHYQQNTPIGDPVLLPDNYLSTQALSKDNPNEKRDHMVLLEFVTAAGITLGMDELYK