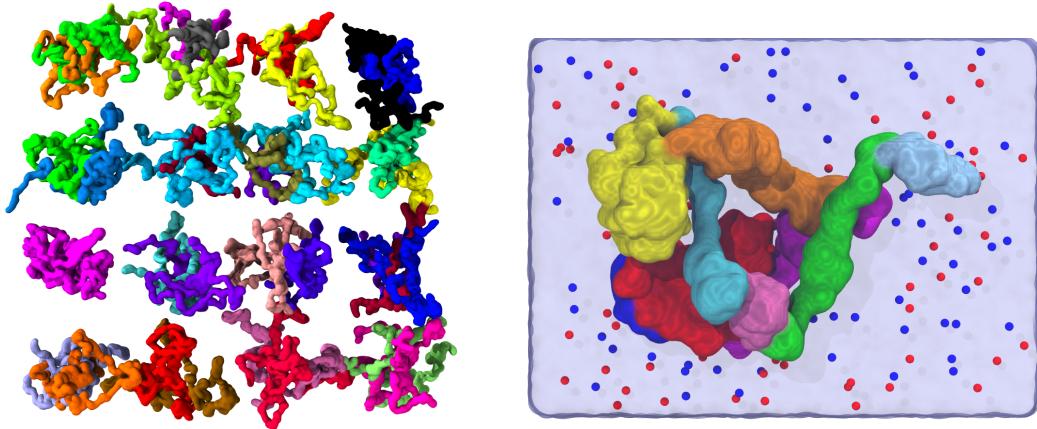


GPU Computing for Bionanotechnology

Mentor: Prof. Aleksei Aksimentiev

Mentee: Swan Htun (SPIN)

July 23 2019



Contents

1 Abstract	3
2 Introduction	3
2.1 Fused in Sarcoma (FUS)	3
3 Methodology	4
3.1 Software & Applications Used	4
3.2 Simulation conditions	4
4 Results	5
5 Discussion	6
6 Conclusion	8
7 Future Directions	8

1 Abstract

Biomolecular condensates are membraneless organelles found in Eukaryotic cells that are formed by liquid-liquid phase separation (LLPS) and display liquid-like properties [5]. An important constituent of certain condensates is the Fused in Sarcoma (FUS) protein, which is an RNA-binding protein involved in various cellular processes such as DNA repair, RNA transport and damage response [5]. The FUS protein is essential for the condensate to undergo the LLPS process in vivo and the FUS protein itself undergoes LLPS in vitro [8]. Our current goal is to computationally investigate phase separation in the FUS protein by studying the phase separation process of both wild-type (WT) and mutant FUS protein under various temperatures and concentrations. In this paper, we present the results of the Brownian Dynamic simulations run at various temperatures ranging from 295K to 450K. The results suggest that the cutoff temperature for phase separation is around 415K. Further simulations at different concentrations are being carried out to quantify the effects of concentration in phase separation process as well as how certain mutations interfere with phase transitions of FUS.

2 Introduction

Biological cells have compartmentalized cellular space for efficient regulation of biochemical reactions by isolating certain biochemical process from the rest of the cell by bounding the reactions inside membranous organelles. Eukaryotic cells, however, also have many organelles that lack membrane-enclosed structures [4], such as stress granules, and nucleoli [4]. Such membraneless organelles are commonly known as biomolecular condensates [3]. Due to their highly dynamic nature, they have liquid-like properties and are formed via liquid-liquid phase separation (LLPS) process as a response to specific biochemical signals [3]. Such properties include the rapid exchange of their constituent molecules with the surrounding cytoplasm or nucleoplasm [4] as well as the ability to deform and freely flow around surfaces of other intracellular structures while under shear stress [3]. They also undergo fusion and fission [5, 4, 3].

These condensates have specific macromolecular composition that can be classified qualitatively into two types: clients, which are preferentially recruited as part of the condensate, and scaffolds, which are required for the formation of the condensate. The majority of the condensate is comprised of clients [3].

The properties and molecular mechanism of these biological condensates have become an interesting topic of research, specifically intrinsically disordered proteins like FUS, as growing evidence suggests their implication in pathogenesis of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) [2]. The physical process involved in biomolecular condensate formation (i.e, phase separation) and certain biomolecules involved are both known, but the precise physical and molecular mechanism of formation and its related processes are still unknown. Thus, it's still a challenge to relate specific features and properties of the condensates with the corresponding biological function, which is essential to understanding how aberrations in them contribute to disease.

2.1 Fused in Sarcoma (FUS)

Each individual FUS protein is made up of 526 amino acids and consists of both ordered and disordered regions that can be divided into a prion-like domain and the RNA binding domain [6]. The phase separation process has been experimentally determined to be driven primarily by the interactions between tyrosine residues from prion-like domains (PLDs) and arginine residues from RNA-binding domains (RBDs) [7]. The PLDs consist mainly of intrinsically disordered regions (IDRs) that are low in amino acid diversity (polar residues and aromatic residues) [7]. Such domains are highly prone to aggregation and have been associated with pathogenesis of diseases such as ALS. Mutations in PLDs have been found to accelerate the conversion of the liquid-like droplets of FUS to solid-state, which is detrimental [5].

3 Methodology

Aspects of phase separation process are simulated using all-atom Molecular Dynamics (MD) and Atomic Resolution Brownian Dynamics (ARBD) simulations. The molecular events involved in phase separation occur at a time scale too fast to resolve experimentally, which is one of the advantages of using computational methods to investigate LLPS. MD numerically solves Newton's equations of motion for each atom in the system to determine the position at each time step and therefore provides atomic-level resolution. ARBD uses Langevin equations of motion to obtain the trajectory of the atoms and does not provide atomic-level resolution since the system must be coarse-grained. The raw output data from both simulations gives the trajectory of all atoms (or coarse-grained particles) of the system. The trajectory is then analyzed to gain insight into the phase separation process. VMD is used to visually aid in the study of LLPS process of FUS.

3.1 Software & Applications Used

ARBD is a GPU-enabled code that takes advantage of the processing power of the GPUs to facilitate fast simulations and achieves better computational efficiency for large systems in comparison to all-atom MD simulations. Python script is used to interact with ARBD. Since the simulation of phase separation phenomena happens at a long time scale, we mainly utilize the BD simulations. Visualizations and the all-atom model of FUS are built using the VMD software. All-atom MD simulations are run using NAMD software on Stampede2 supercomputer.

We began by building a complete all-atom model of FUS protein using publicly available data of the atomic coordinates and the structural information of each ordered segment of FUS protein on Protein Data Bank: <http://www.rcsb.org>. As no structures are available for the disordered regions, their structures are generated from the amino acid sequence using the Avogadro software. Then, each individual segments are linked using VMD. From the all-atom model, we build the CG model of FUS as seen in Fig. 1 using the python.

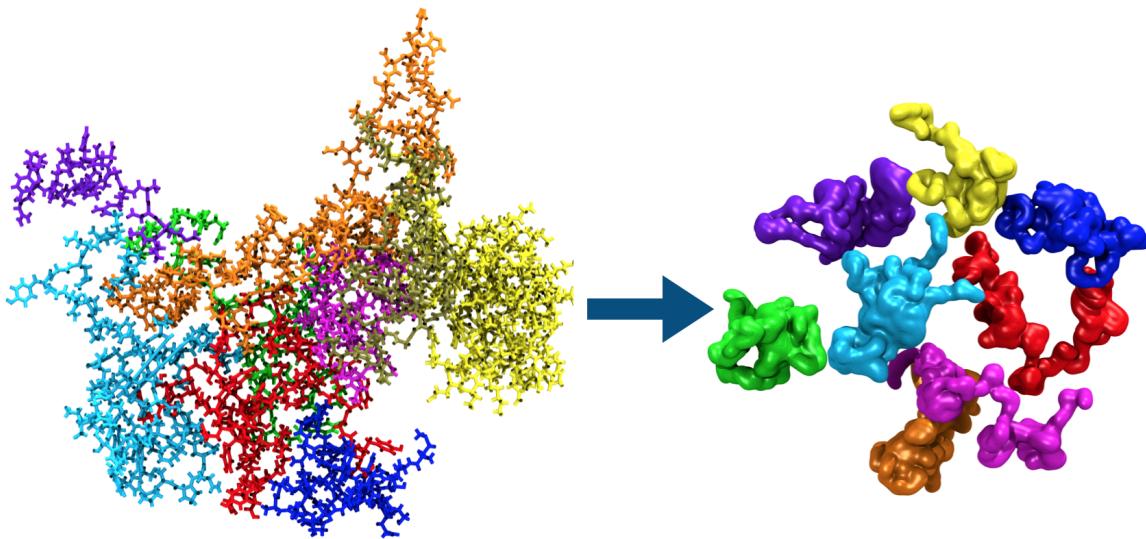


Figure 1: All-atom model represented in ‘bonds’ presentation (left) and coarse-grain model represented using ‘quicksurf’ (right) of FUS using VMD. As shown, there are 1 protein in all-atom system and 8 proteins in the CG system.

3.2 Simulation conditions

All-atom simulations of WT FUS protein are run at 300K with NPT conditions. The ARBD simulations of 8 WT FUS proteins are run at 295K, 350K, 400K, 410K, 420K, 430K, 440K and 450K in NVT conditions for over 10 μ s. Another set of ARBD simulations of 64 WT FUS proteins as well as the mutant FUS proteins at similar temperatures are also being run to reveal the impact of concentration and mutations on phase separation. We also ran all-atom simulation of FUS

protein for 100ns. It's important to know the concentration $[FUS]$ of FUS in the system, which can be calculated via

$$[FUS] = \frac{n}{N_A} \cdot \frac{1}{V}$$

where n is the number of FUS proteins, N_A is Avogadro's number, and V is the volume of simulation box.

4 Results

When looking at the simulated trajectory of the FUS protein over 8 μs using VMD, we can see that at temperatures lower than 415K, proteins coalesce to form a cluster or a few clusters as seen in Fig. 2 (b). We created density maps of the protein densities to see the density distribution of the system Fig.4. We should expect the more homogeneous system, the one that did not coalesce to have a more uniform density map. Then, FFT is applied to the density map to quantify the phase transition from homogeneous to inhomogeneous phase, Fig 5(b). The normalized density fluctuations are analyzed to observe the temperature at which phase separation occurs.

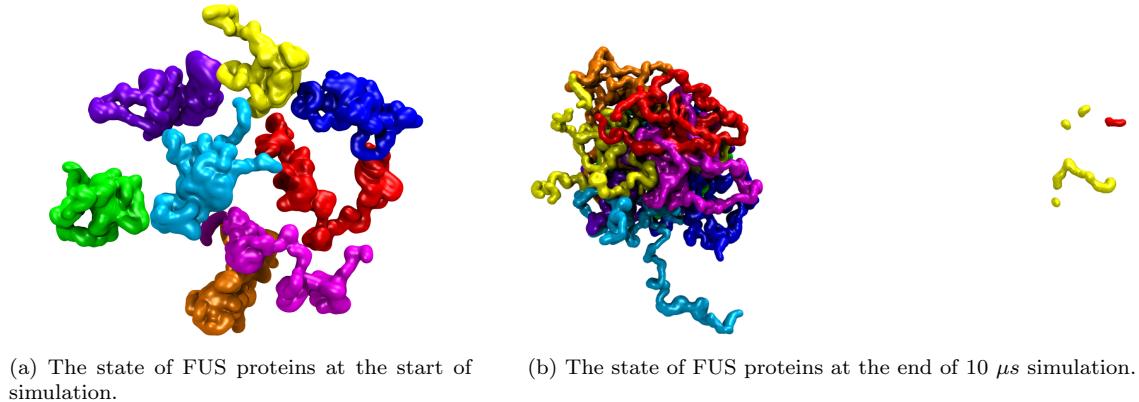


Figure 2: ARBD Simulation of FUS proteins. 8 FUS proteins were run at 350K in a 825 \AA^3 resulting in $23.65 \mu M$. The total number of atoms present were 43444.

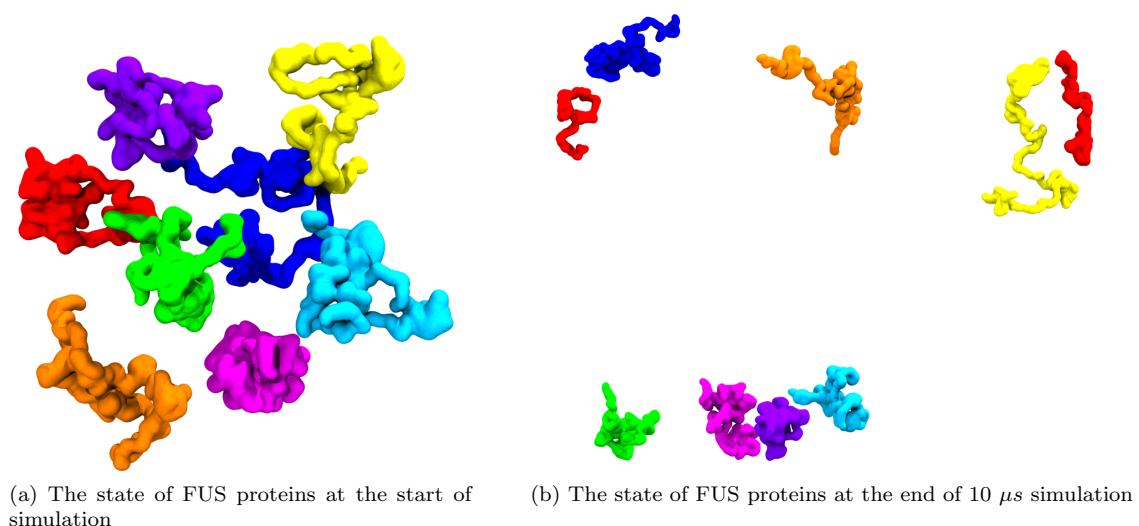


Figure 3: ARBD Simulation of FUS proteins. 8 FUS proteins were run at 450K in a 825 \AA^3 resulting in $23.65 \mu M$. The total number of atoms present were 43444.

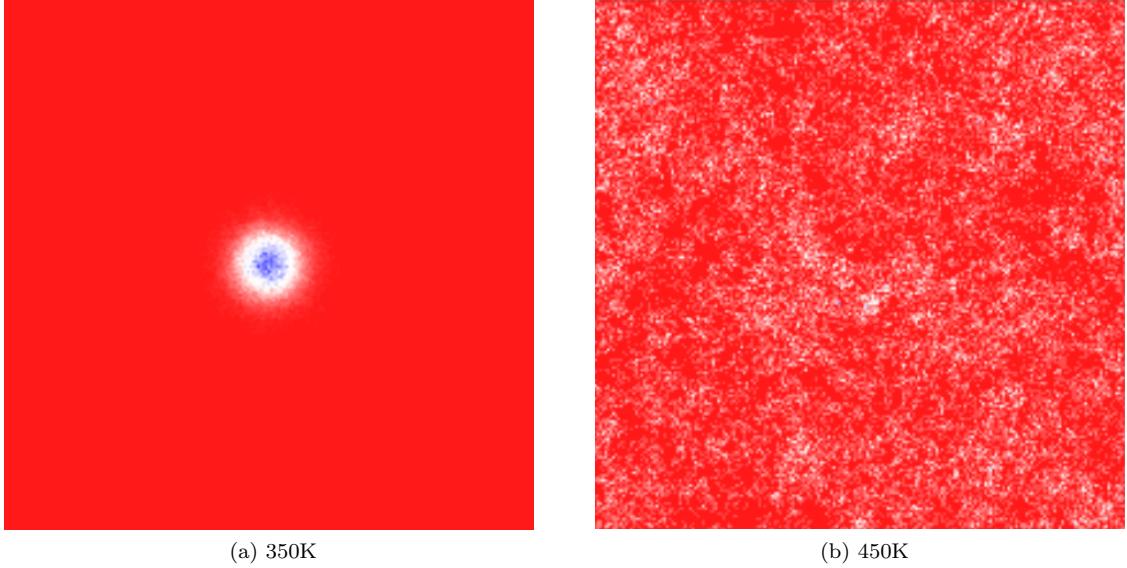


Figure 4: Density maps of FUS proteins obtained using the data from ARBD simulation

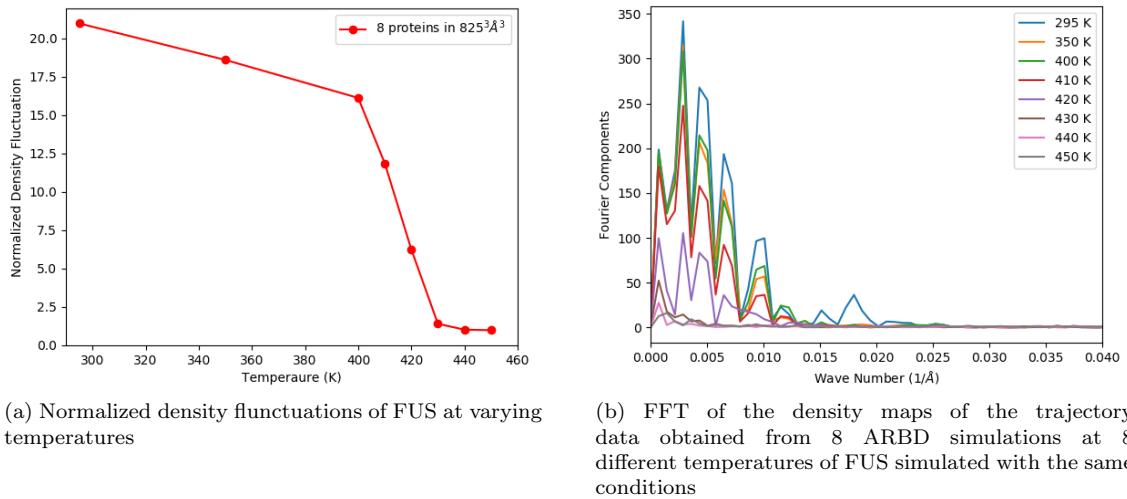


Figure 5: Transition graph and FFT of the density of the FUS proteins at varying temperatures. 8 FUS proteins were run at various temperatures in 825\AA^3 resulting in $23.65 \mu\text{M}$. The total number of atoms present were 43444.

5 Discussion

The results of proteins undergoing liquid-liquid phase separation resembles liquid droplets, which is caused by the condensation into a dense phase [1]. LLPS is driven by the exchange of macromolecule/water interactions for macromolecule/macromolecule and water/water interactions under conditions for which this process is energetically favorable [1]. Therefore, the LLPS process is sensitive to environmental conditions such as temperature and concentration also plays an important role.

To investigate the temperature dependence of the LLPS process of FUS, 8 simulations are run for $10 \mu\text{s}$. The state of the system at the end of $10 \mu\text{s}$ ARBD simulation ran at 350K appears to be clustered Fig.2(b). In contrast, the system which was run at 450K, Fig.3(b), we can see that the proteins have not coalesced at all. By creating density maps of these, Fig.4, we can see that the distribution of the density map supports the results seen in Fig.2(b) and 3(b). For the system that didn't coalesce Fig.4(b), the distribution of the density is more uniform unlike the other system, where we see a big dense region in the density map Fig.4(a) reflecting the cluster in 2(b).

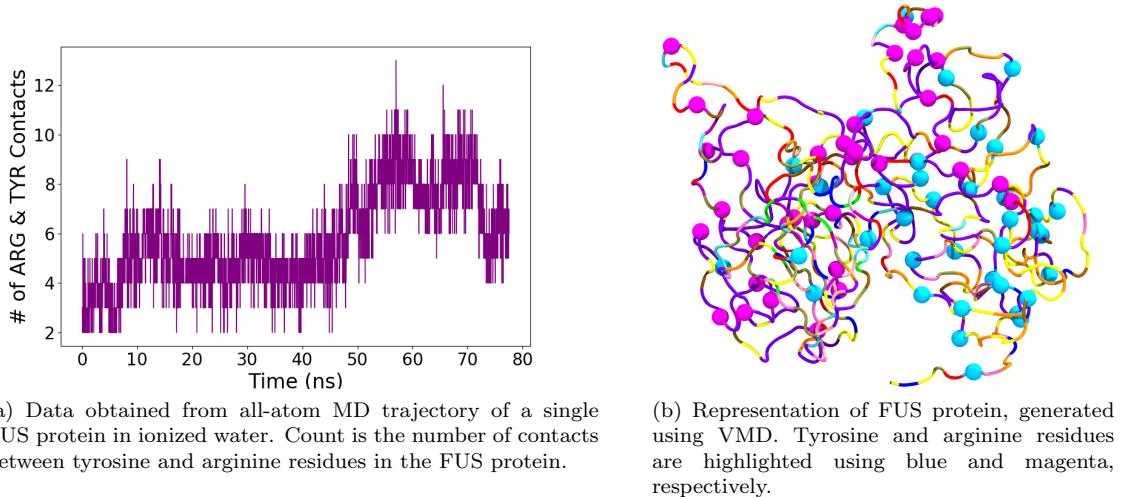


Figure 6: Investigations of tyrosine and arginine residues in single FUS protein.

Among the 10 simulations run at aforementioned temperatures for the $23.65 \mu M$ system, around 415K, we can see a transition from inhomogeneous to homogeneous phase in Fig.5(a). We can clearly see that the normalized density fluctuations decreases as temperature increases indicating that for temperatures above 420K, the density of the system remained almost the same as before. We can also observe an abrupt drop in normalized density fluctuations around 415K.

Fig.5(b) shows the FFT of the density map against the wavenumber for systems at varying temperatures. If the system is inhomogeneous, there are more density fluctuations yielding closely spaced wave lengths and hence, a large wavenumber. On the other hand, if the system is homogeneous, the density fluctuations are more uniform resulting in larger wave lengths and hence, a small wavenumber. If the system is inhomogeneous, we would see more peaks at larger wavenumbers as well as see peaks of larger magnitude. For a homogenous system, we would see very few peaks, if any at small wavenumber. We see an example of a homogenous system at 440K, indicating that a condensate has not formed. Comparing the plot for 440K and the plot for 295K, we see many more peaks at 295K indicating an inhomogeneous phase as phase separation occurs. From Fig.5(b), we see less homogeneity in the 295K based on a smaller section of the data being at the zero line.

From previous studies [7], we know that interactions between tyrosine and arginine residues are the principal drivers of LLPS, so one of the analysis includes counting the total number of contacts between the tyrosine and arginine residues over the 80ns simulation, Fig.6(a).

Each simulation in this study took approximately 24 hrs to run when 1 unit of GPU is used for $23.65 \mu M$ system. To understand how concentration affects the phase separation process, higher concentration systems are being simulated at the same temperatures. However, these simulations take almost 3 times longer than the $23.65 \mu M$ system so getting the results requires more time. The all-atom simulations take even longer than ARBD - they take about a week to simulate 100 ns. It is likely that there maybe ways to optimize the system to gain better computational run time. Given the duration of the internship and the availability of GPUs, we are limited in how many simulations we can run. Moreover, the process of writing code to analyze the MD and ARBD trajectories adds even more time. Even usage of supercomputers still have time constraints as there is a limit on how many node hours one job submission will have access per day.

6 Conclusion

Preliminary studies on interactions between tyrosine and arginine residues within a FUS protein were conducted using all-atom MD. The liquid-liquid phase separation process was studied using coarse-grained ARBD simulations. The liquid-liquid phase separation process is highly temperature dependent as at temperatures higher than 420K, the proteins didn't coalesce at all. However, at room temperature or temperatures lower than 420K, the proteins aggregate. Hence, the threshold temperature at which LLPS occurs appears to be at 420K (see Fig.5(a)). The VMD representations of the process as well as the analysis of the ARBD results previously presented all support this conclusion. All simulations conducted in this paper were done using the WT FUS protein.

7 Future Directions

Additional simulations at various concentrations of FUS protein should be run to find out the concentration threshold for the LLPS process, akin to how we determined the temperature threshold in this paper. Another promising direction is to compare the relevant behavior of the WT FUS protein with that of the mutant FUS protein, which requires that we run analogous mutant FUS proteins simulations. The results of this comparison would provide insight into how certain mutations affect the phase separation properties of FUS. Moreover, it would help us determine exactly which regions of the protein are most involved in LLPS, for mutations in an essential region should disrupt the LLPS process much more than mutations in nonessential regions.

References

- [1] Simon Alberti, Amy Gladfelter, and Tanja Mittag. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, 176:419–434, 01 2019.
- [2] Hao Deng, Kai Gao, and Joseph Jankovic. The role of fus gene variants in neurodegenerative diseases. *Nature reviews. Neurology*, 10, 05 2014.
- [3] Salman F Banani, Hyun Lee, Anthony A Hyman, and Michael K Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews. Molecular cell biology*, 18, 02 2017.
- [4] A. Mullard. Biomolecular condensates pique drug discovery curiosity. *Nature Reviews Drug Discovery*, 2019.
- [5] Avinash Patel, Hyun Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M Franzmann, Andrei Pozniakovsky, Ina Poser, Nicola Maghelli, Loic Royer, Martin Weigert, Eugene W Myers, Stephan Grill, David Drechsel, Anthony Hyman, and Simon Alberti. A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell*, 162:1066–1077, 08 2015.
- [6] Dylan T. Murray, Masato Kato, Yi Lin, Kent Thurber, Ivan Hung, Steven McKnight, and Robert Tycko. Structure of fus protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell*, 171, 09 2017.
- [7] Jie Wang, Jeong-Mo Choi, Alex Holehouse, Hyun Lee, Xiaojie Zhang, Marcus Jahnel, Shovamayee Maharana, Régis Lemaitre, Andrei Pozniakovsky, David Drechsel, Ina Poser, Rohit Pappu, Simon Alberti, and Anthony Hyman. A molecular grammar governing the driving forces for phase separation of prion-like rna binding proteins. *Cell*, 174, 06 2018.
- [8] Takuya Yoshizawa, Rustam Ali, Jenny Jiou, Ho Yee Joyce Fung, Kathleen Burke, Seung Joong Kim, Yuan Lin, William B Peebles, Daniel Saltzberg, Michael Soniat, Jordan Baumhardt, Rudolf Oldenbourg, Andrej Sali, Nicolas Fawzi, Michael K Rosen, and Yuh Min Chook. Nuclear import receptor inhibits phase separation of fus through binding to multiple sites. *Cell*, 173, 04 2018.