

INTRA-MOLECULAR PROTEIN DYNAMICS IN LIQUID-LIQUID
PHASE SEPARATION

by

Katukurunde Kasun Gayasri Gamage

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Physics – Doctor of Philosophy

2025

Abstract

To investigate protein behavior *in vitro* it is important to mimic the cellular environment. Given the high densities of ~ 400 mg/mL and the large number of a variety of macromolecules present in cells, reproducing this complex crowded environment *in vitro* has been a difficult task. Most of the time crowding is achieved through synthetic polymers such as polyethylene glycol (PEG) or concentrating the protein of interest itself which can underestimate interactions and hide the true protein behavior. We expect condensates formed through liquid-liquid phase separation (LLPS) to be a better platform to study proteins or RNA such as scaffold-client condensate systems which can incorporate a variety of macromolecules under physiological conditions. In this work we study such condensate environments and investigate protein behavior.

We investigated a short polymer system of the RNA poly-adenine with the peptide RGRGG to understand the length dependence of LLPS and its underlying thermodynamics. Results showed that the length dependence of LLPS is primarily driven by the entropy of confinement. To study macromolecular behavior *in vitro* using condensates as a crowded platform, we introduce Trp-Cys quenching as a technique to measure the dynamics of intrinsically disordered proteins (IDPs) in condensates. We were able to successfully calculate the intra-molecular diffusion of α -synuclein in RLP condensates, and the protein appeared to be highly dynamic in the condensed phase showing only ~ 50 % slowdown in intra-molecular diffusion compared to its monomeric state in the dilute solution even though the concentration in condensates was ~ 300 times high. We also developed the Trp-Cys quenching technique to characterize transient inter-molecular interactions of proteins under crowded conditions. Using Villin variants and drkN SH3 domain with Protein G as the crowder the technique was able to capture the difference between long-range repulsive and attractive interactions, the strength of short-range interactions, concentration variations and interaction preference for different sites. Furthermore, we investigated the early-stage aggregation of O-GlcNAcylated α -synuclein at two different sites T72 and S87. Results indicated that while glycosylation at T72 can suppress early-stage aggregation S87 can promote it. We were able to hypothesize that this distinction is because the most common transient interactions for these two sites are different.

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my supervisor, Dr. Lisa Lapidus, for her unwavering support and insightful critiques throughout my research journey. She had been an incredible mentor, extremely kind, open-minded, and her guidance has helped me accomplish my academic goals with great success.

I am also grateful for the support given by Dr. Michael Feig, especially providing me with knowledge and guidance on computational work. I am thankful to my committee members Dr. Matthew Comstock, Dr. Norman Birge and Dr. Remco Zegers for providing me with constructive feedback on my research projects and presentations.

I am grateful for the help that I got from all my lab members to make this work a success. Special thanks go to Joyce Goodluck for mentoring me when I started in the lab, Dr. Gilberto Valdes-Garcia and Dr. Alexander Jussupow for their help with computer simulations and to Thong Van for his support in the lab. I would also like to thank Dr. Melinda Frame for the help she gave me with Confocal imaging.

I would like to thank my loving wife, who has been with me throughout this journey always pushing me to do better and my parents who have been extremely supportive throughout my entire life, always emphasizing the importance of education by word and example. Their encouragement has been immense. I am also very grateful for the support of in-laws, and other family members.

Finally, I would like to recognize and thank all others who have collaborated in this work and not mentioned here, Michigan State University for the support given and the National Science Foundation for the funding provided.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	1
1.1: Liquid-liquid phase separation.....	2
1.2: Protein dynamics.....	4
CHAPTER 2: Methods.....	7
2.1: Trp-Cys quenching technique.....	7
2.2: SSS theory.....	9
2.3: Pump-probe spectroscopy of Trp-Cys quenching.....	9
2.4: Syringe pump system.....	11
2.5: Coarse-grained simulations.....	13
2.6: Expression of α -syn and RLP.....	15
2.7: Fluorescent labeling of α -syn and RLP.....	17
2.8: Dynamic light scattering of condensates.....	19
2.9: Isothermal titration calorimetry of condensate formation.....	20
CHAPTER 3: Length dependance of liquid-liquid phase separation.....	22
3.1: Phase separation of the peptide-RNA system.....	22
3.2: The phase boundary.....	26
3.3: Condensate growth.....	27
3.4: Liquid nature of the condensates.....	29
3.5: Induction of LLPS by short peptides.....	31
3.6: Enthalpy of phase separation.....	32
3.7: Confinement entropy.....	35
3.8: Free energy of phase separation.....	36
CHAPTER 4: RLP condensate system.....	38
4.1: RLP sample preparation.....	38
4.2: Optimizing conditions for Trp-Cys measurements.....	40
4.3: Partitioning of α -syn in the condensate solution.....	42
CHAPTER 5: Intra-molecular diffusion of α -syn in condensates.....	45
5.1: α -synuclein dynamics under dilute conditions.....	45
5.2: Trp-Cys distance distribution in condensates.....	47
5.3: Inter-molecular quenching in condensates.....	49
5.4: Translational diffusion of α -syn in condensates.....	50
5.5: Intra-molecular diffusion of α -syn in condensates.....	52
CHAPTER 6: Characterization of protein-protein interactions in a crowded system.....	56
6.1: Trp-Cys protein-protein quenching.....	56
6.2: Trp-Cys interaction potential.....	59
6.3: Survival probability.....	60
6.4: Surface preference for interactions	63
CHAPTER 7: Dynamics and early-stage aggregation of O-GlcNAcylated α -synuclein.....	64
7.1: Fibrilization of O-GlcNAcylated α -synuclein.....	64
7.2: Intra-molecular diffusion.....	65
7.3: The kinetic model of aggregation.....	67

CHAPTER 8: Conclusion.....	72
BIBLIOGRAPHY.....	75
APPENDIX.....	79

LIST OF TABLES

Table 2.1: Measured viscosities of the samples for each sucrose percentage at 37°C.....	12
Table 2.2: Bonded and non-bonded interaction parameters used in COCOCMO and COCOMO2.....	14
Table 5.1: Diffusion coefficient in the condensed phase for the α -syn mutants.....	54
Table 6.1: Amino acid sequences and electrostatic charge of Villin variants, SH3 and Protein G.....	57
Table 7.1: Computed rates and diffusion coefficients of α -syn, α -syn(gT72) and α -syn(gS87) at 37°C.....	67
Table A1: Concentrations of PolyA and RGRGG used for ITC.....	90
Table A2: LLPS at different concentrations of PolyA and RGRGG.....	91
Table A3: Residue-specific parameters used in COCOMO, COCOMO2 and COCOMO1.2 σ simulations.....	91
Table A4: Potential parameters used in Monte Carlo simulations to match atomistic Probability distributions.....	92

LIST OF FIGURES

Figure 2.1: Kinetic model used for the Trp-Cys quenching process.....	7
Figure 2.2: Schematic of the pump-probe spectroscopy setup.....	10
Figure 2.3: Background Trp quenching due to the non-degassed volume.....	11
Figure 2.4: Schematic of pump-probe spectroscopy coupled with a syringe pump system used for injecting protein samples for Trp-Cys measurements.....	12
Figure 2.5: Schematic outlining the DLS technique.....	19
Figure 2.6: ITC setup and recorded heat measurements.....	20
Figure 3.1: Length-dependent LLPS of PolyA-RGRGG mixtures using COCOMO.....	23
Figure 3.2: Experimental and COCOMO 1.2 σ results for LLPS of PolyA-RGRGG mixtures.....	24
Figure 3.3: Cluster density of condensates and phases in the LLPS system.....	25
Figure 3.4: Particle size of PolyA-RGRGG mixtures measured by DLS.....	27
Figure 3.5: Size distribution analysis of PolyA-RGRGG condensates over time.....	28
Figure 3.6: Liquid behavior of PolyA-RGRGG condensates.....	29
Figure 3.7: Phase separation recovery of PolyA-RGRGG condensates by short peptides.....	31
Figure 3.8: ITC binding isotherms for various mixtures of PolyA and RGRGG.....	32
Figure 3.9: Phase separation boundary of PolyA and RGRGG mixtures at 0.1mg/ml.....	33
Figure 3.10: Pairwise RDF between different PolyA and RGRGG residue types in the condensed phase.....	34
Figure 3.11: Energetic analysis of PolyA-RGRGG phase separation.....	36
Figure 4.1: Removal of condensate clusters from RLP samples.....	39
Figure 4.2: Incorporation of α -syn in RLP condensates.....	40
Figure 4.3: Dependence of RLP optical density on temperature and RLP concentration.....	40
Figure 4.4: C_{sat} of RLP from confocal microscopy.....	41
Figure 4.5: Diagram of dilute vs condensate samples during Trp-Cys measurements.....	42
Figure 4.6: Calculating the partitioning of α -syn in a condensate sample using confocal microscopy.....	43
Figure 4.7: Confocal images of α -syn with and without RLP captured along the z-axis.....	44
Figure 4.8: Partition coefficients of α -syn and RLP in condensates.....	44
Figure 5.1: Diffusion analysis of α -syn in dilute solution.....	46
Figure 5.2: C_{sat} of RLP for different starting conditions in CG simulation.....	48

Figure 5.3: Trp-Cys probability of α -syn in the condensed phase.....	48
Figure 5.4: Inter-molecular quenching of α -syn in the condensed phase.....	49
Figure 5.5: FRAP of α -syn ^{39w69c} in RLP condensates.....	51
Figure 5.6: High sucrose concentrations dissolve α -syn-RLP condensates.....	52
Figure 5.7: Trp decay rates in dilute vs condensate solutions.....	53
Figure 5.8: Intra-molecular diffusion coefficients of α -syn mutants in the condensed phase.....	53
Figure 5.9: Residue-residue contact frequencies between proteins in α -syn-RLP condensate systems.....	55
Figure 6.1: Kinetic model for the Villin/SH3-Protein G crowded system	56
Figure 6.2: Experimental measurements of Trp triplet lifetime in Villin variants and SH3 in the presence of Protein G.....	58
Figure 6.3: 1D Monte Carlo potential analysis of Villin WT-Protein G system.....	59
Figure 6.4: Comparing atomistic survival probabilities to experimental Trp decays for Villin/SH3-Protein G systems.....	61
Figure 6.5: Atomistic interaction potentials for the Villin/SH3-Protein G systems.....	62
Figure 6.6: Residue-wise contacts between Protein G and Villin variants.....	63
Figure 7.1: Fibrilization kinetics of glycosylated α -syn	64
Figure 7.2: Measured decay rates of glycosylated α -syn as a function of viscosity.....	65
Figure 7.3: Computed rates and diffusion coefficients of glycosylated α -syn.....	66
Figure 7.4: Kinetic model of aggregation for glycosylated α -syn.....	70
Figure 7.5: Site dependent reconfiguration model.....	71
Figure A1: Inducement of condensates by Cy3 in PolyA ₁₀ [RGRGG] ₂	79
Figure A2: Radii of gyration and distance distributions for PolyA-RGRGG mixtures.....	84
Figure A3: Trp decay transition from dilute to condensed phase in α -syn-RLP system.....	86
Figure A4: Survival probability as a function of quencher concentration from CG simulations.....	87
Figure A5: The effect of system size on PolyA _N -[RGRGG] _M LLPS simulations.....	88
Figure A6: Particle size distributions of PolyA-RGRGG mixtures measured by DLS.....	89
Figure A7: Starting conditions and equilibration of the α -syn-RLP CG systems.....	89

Chapter 1

Introduction

Properties of proteins and nucleic acids have been extensively studied *in vitro* under dilute conditions. Experiments are typically carried out under physiological conditions of pH and temperature to reproduce cellular behavior. Previous work in our lab on the intra-molecular dynamics of α -synuclein to understand its aggregation propensity is a good example. These studies were targeted on investigating familial and aggregation inhibitory mutations relative to the unmutated protein[3]. But the true behavior of the protein in its cellular environment is not extensively captured. Therefore, it is important to mimic the cellular environment *in vitro* experiments. One perilous task is mimicking its high density.

Unlike in a dilute solution cells are crowded with high concentrations of proteins, nucleic acids and other macromolecules. The density of the cytosol has been estimated to be as high as 400 mg/mL consisting of thousands of different types of macromolecules. This complex environment is not easily achieved *in vitro*. Abundantly used crowders *in vitro* experiments are PEG, Ficoll, dextran and sucrose which are charge-neutral polymers therefore interactions found in a cellular environment are underestimated. *In vitro* crowding is also achieved at high concentrations of the protein itself. Again, this will not mimic the variety of macromolecules found in cells. Achieving high concentrations of biomolecules *in vitro* encounters solubility limits in buffered solutions leading to the usage of organic solvents such as dimethyl sulfoxide (DMSO). These factors limit the number of macromolecular systems that can be studied under physiologically relevant crowded conditions. We expect condensates to be a better platform to study a vast range of proteins or nucleic acids under crowded conditions.

Condensates are highly concentrated membrane-less compartments formed by macromolecules such as proteins and nucleic acids, generally through LLPS. Many types of biomolecular condensates are found in cells and are involved in various cellular activities. And many types of biomolecules have been observed to form condensates *in vitro* mainly driven by electrostatic interactions and possess concentrations as high as 200 - 400 mg/mL in the condensed phase even in the absence of a crowder[4-6]. Other biomolecules have been shown to form condensates and act as scaffolds to recruit many other types of proteins to be studied in the crowded environment

of the host molecule[7]. Synthetic crowders such as PEG are also used to condense a variety of proteins or nucleic acids together where they encounter multiple types of interactions[8]. In this study we recruit α -synuclein (α -syn) as a client into resin-like polypeptide (RLP) condensates which act as the scaffold to study the dynamics of α -syn in a crowded environment.

Here we develop the Trp-Cys quenching, a pump-probe spectroscopy technique typically used in measuring intra-molecular dynamics of disordered polypeptide chains, to measure protein dynamics in condensates using bulk samples and to characterize protein-protein interactions of folded proteins in crowded systems. In addition, we investigate the underlying thermodynamics of a length-dependent LLPS system. The dynamics and the aggregation propensity of the Parkinson's disease causing IDP α -syn is investigated under crowded or dilute conditions through this study as well.

1.1 Liquid-liquid phase separation

LLPS is a mechanism by which certain proteins and nucleic acids encounter to form a liquid-like globular dense phase that can coexist with its dilute phase. Through phase separation, biological systems selectively compartmentalize proteins and nucleic acids creating a highly concentrated environment that can increase reaction kinetics and can isolate certain macromolecules to prevent pathogenic encounters.

Many such liquid-like membrane-less compartments are found in cells, such as nucleoli[9], paraspeckles[10], stress granules[11] and germ granules[12] which are known to regulate cellular activity such as processing and modification of RNA, post-translational modification of proteins, gene regulation, responding to alterations in the cellular environment such as stress and cell development. For example, P granules in germ cells are condensates formed of many proteins and RNA. It has been observed that embryonic division in *Caenorhabditis elegans* is a result of the formation of P granules and its asymmetric localization which is driven by a polarity protein gradient that controls the saturation conditions of phase separation[12]. In another example, formation of stress granules by phase separation of the RNA binding protein hnRNPA1 has been observed with induced stress *in vivo* and *in vitro* [11, 13, 14]. Studies indicate phase separation leading to fibrilization with pathogenic mutations which are rescued by point mutations, resulting

in dissolving away the condensates with relaxation of stress. Another interesting condensate system observed is the phase separation of α -syn with synapsin 1 in neurons[15] where synapsin 1 is essential for the phase separation of α -syn and it is enhanced with the presence of synaptic vesicles.

Condensates are often formed under high concentrations, where at least one type of polymer acts as a crowder driving the system to phase-separate. Nucleation of condensates typically occurs at a threshold concentration above which the system goes to saturation. This threshold can be adjusted by external factors such as the temperature, pH, and salt content. In a dilute solution macromolecules interact through fewer and weak transient interactions with each other and with the solvent molecules, specifically water. Near threshold concentrations macromolecules have the tendency to interact favorably and phase-separate in instances where interactions between macromolecules are stronger than with water. In such systems it is understood that inter- and intra-molecular transient interactions between macromolecules balance against entropy driving the system towards thermodynamic equilibrium, hence minimizing the free-energy compared to a highly concentrated dilute phase[16-18]. LLPS exhibits coexistence between the dilute and the condensed phase by maintaining the same chemical potential in both the phases but at different concentrations. This allows molecules to freely move between phases while keeping the system at equilibrium. Transient electrostatic, or aromatic interactions maintain the liquid-nature of condensates. In addition to weak transient interactions, strong and reversible interactions have been observed through post-translational modifications[19]. The abundance of interacting domains or motifs in macromolecules is a key factor in LLPS and multivalency observed in proteins and RNA in the form of either repeat units or different types of domains seems to promote phase separation[20]. Several forms of nucleation of condensates have been observed. For example, in the scaffold-client system, one type of protein forms condensates creating a scaffold for the client protein to partition into[21]. PML nuclear bodies[22] and P bodies[23] are some examples *in vivo*, while resin-like polypeptides[24] have been synthesized to work as scaffold in *in vitro* experiments. A more common nucleation mechanism is the co-condensation of multiple proteins starting from their monomeric state such as α -syn and synapsin 1 in neurons[15]. Another example is oppositely charged peptides and nucleotides such as different lengths of Lysine and Arginine with ATP (Adenosine triphosphate) or ADP (Adenosine diphosphate)[25]. A third method is the

driving of a protein to form condensates by a molecular crowder, typically observed in *in vitro* experiments where a crowder polymer is used to mimic the highly concentrated environment of a cell. One of the most common crowder used is PEG, which drives IDPs such as α -syn to phase-separate[26].

To get a good understanding of the mechanisms underlying LLPS and the functions of condensates, properties such as protein dynamics, thermodynamics, molecular interactions, response to temperature and concentration need to be studied in detail.

1.2 Protein dynamics

The condensed environment in LLPS shows distinct properties relative to its dilute counterpart mainly due to its density. Properties such as viscosity and molecular density are increased while molecular dynamics has been observed to slow down. Various microscopy and spectroscopy techniques have been implemented to analyze protein dynamics. Translational dynamics of proteins in condensates have been extensively studied through techniques such as fluorescence recovery after photo-bleaching (FRAP) and fluorescence correlation spectroscopy (FCS). FRAP data for folded as well as intrinsically disordered proteins *in vivo* and *in vitro* condensates shows apparent translational diffusion coefficients ranging from $0.0001 - 0.1 \text{ A}^2\text{ns}^{-1}$ [4, 12, 27-29]. Translational diffusion coefficients measured through FCS are typically two orders of magnitude higher than FRAP measurements[4, 30]. FCS is considered more accurate than FRAP due to its single-molecular resolution. Inter- and intra-molecular dynamics of proteins have been studied using single-molecular Förster resonance energy transfer (smFRET). FRET transfer efficiency and reconfiguration time measurements have been analyzed to understand protein conformations and dynamics in condensates respectively[4, 31-33]. Dynamics around an order of magnitude slower have been observed in the condensed phase relative to the dilute phase monomers and third of a fraction slower than dimers. These studies have indicated that concentration and viscosity increase by ~ 1000 and ~ 300 times with phase-separation respectively.

The Trp-Cys quenching technique (**Section 2.1**) which is used in this study has been typically implemented to investigate the intra-molecular dynamics of disordered proteins in dilute solutions[34-38]. It is a spectroscopic technique that uses pump-probe spectroscopy (**Section 2.3**)

to perform bulk measurements on protein samples. It is sensitive in capturing nano to microsecond dynamics and exhibits sub-nanometer length resolution due to the distance dependent Trp-Cys quenching (**Eq 2.1**) measurements carried out at close residue-residue contact. This makes it suited for investigating transient protein-protein interactions, even under crowded conditions. A further advantage of this technique is that only natural amino acids are involved and additional labeling that may interfere with molecular interactions can be avoided. Earlier work on proteins like α -syn, a protein involved in Parkinson's disease has shown correlation between its intra-molecular dynamics and aggregation propensity. It has been observed that aggregation is kinetically controlled by intra-molecular diffusion, suggesting that the faster the reconfiguration of a protein chain, the lower the probability to stably associate with another protein, thereby reducing the propensity to aggregate. The dynamics of different regions of the α -syn chain have shown to be similar[39]. Familial mutation of α -syn such as A53T has been shown to slow down chain dynamics while artificial mutations such as V74E, which shows fibril elimination in cells, exhibits faster chain dynamics[3, 40]. Incorporating experimental techniques such as FRET and Trp-Cys quenching with molecular dynamics simulations has resulted in acquiring information on protein behavior in dilute and condensate systems typically unavailable through experiments alone. Validating simulation data with factors such as protein diffusion[4] or conformational dynamics[36] has produced information on molecular interaction mechanisms in condensate systems and possible distance dependent diffusion behaviors of proteins under dilute conditions.

In **Chapter 3** a short polymer peptide-RNA system is studied to understand how the length of polymers can affect LLPS and the underlying thermodynamics. Different lengths of the RNA polymer poly-adenine are tested with various repeats of the peptide RGRGG through experimental and computational techniques. **Chapter 4 and 5** introduce Trp-Cys quenching as a technique to measure the intra-molecular diffusion coefficient of an IDP when inside a condensate. Measurements are carried out collectively through pump-probe spectroscopy, confocal microscopy and CG simulations. We incorporate the IDP α -syn into RLP condensates and investigate the dynamics of α -syn in the condensed phase. **Chapter 6** goes through the process of developing the Trp-Cys quenching technique to probe transient protein-protein interactions in crowded protein systems. We investigate protein systems of Villin variants or SH3 with Protein G as the crowder and results are interpreted by molecular dynamic simulations. In **Chapter 7** we investigate the

early-stage aggregation or the small oligomer formation propensity of glycosylated α -syn under dilute conditions for two different sites of glycosylation, T72 and S87 using the Trp-Cys quenching technique.

Chapter 2

Methods

2.1 Trp-Cys quenching technique

The Trp-Cys quenching technique is used to calculate the intra-molecular diffusion of IDPs using pump-probe spectroscopy (**Section 2.3**). The protein chain is mutated at two distinct points with a Trp and a Cys and the Trp is then excited to a long-lived triplet state using a 289 nm ultra-violet (UV) pulsed laser beam. Trp exhibits a triplet state lifetime of $\sim 40 \mu\text{s}$ [35] at room temperature for 15 μM N-acetyl-L-tryptophanamide (NATA). And in the presence of an efficient quencher such as Cys this is reduced upon close contact[35]. The mechanism of quenching is thought to occur via electron transfer from the tryptophan triplet state to the sulfur atoms of the quencher[41-43] and has an exponential dependence on the intra-molecular Trp-Cys distance and is given by:

$$q(r) = q_0 \exp(\beta(r - a)) \quad (\text{Eq. 2.1})$$

where $q_0 = 4.2 \times 10^9 \text{ s}^{-1}$ and $\beta = 4.0 \text{ \AA}^{-1}$ are the amplitude and the decay rate respectively which have been experimentally determined[34]. Here a is the closest distance of approach between the two contacts. The decay of the Trp population due to Cys quenching is then probed using a 445 nm continuous laser.

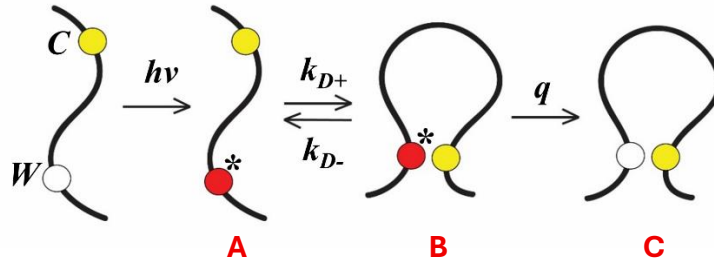


Figure 2.1: Kinetic model for the Trp(W)-Cys(C) quenching process. UV radiation excites the W to a long-lived triplet state and once the W and the C diffuse towards each other at a rate of k_{D+} they will either quench at a rate of q at close contact or diffuse away at a rate of k_{D-} . Excitation is indicated as $*$.

The observed decay (k_{obs}) of the Trp is then kinetically modeled (**Figure 2.1**) where the two contacts can diffuse toward each other at a diffusion-limited rate of k_{D+} and quench at a rate of q or diffuse away from each other at a rate of k_{D-} without quenching.

The kinetic model can be solved as

$$\frac{dA}{dt} = -k_{obs}A = -k_{D+}A + k_{D-}B \quad (Eq. 2.2)$$

$$\frac{dB}{dt} = k_{D+}A - (q + k_{D-})B = 0 \text{ (at steady state)} \quad (Eq. 2.3)$$

$$\frac{dC}{dt} = qB \quad (Eq. 2.4)$$

Solving **Eq. 2.2**, **2.3** and **2.4** gives the observed decay rate as

$$k_{obs} = \frac{k_{D+}q}{k_{D-} + q} \quad (Eq. 2.5)$$

which can be rearranged as

$$\frac{1}{k_{obs}} = \frac{1}{k_R(T)} + \frac{1}{k_{D+}(T, \eta)} \quad (Eq. 2.6)$$

where

$$k_R = q \frac{k_{D+}}{k_{D-}} \quad (Eq. 2.7)$$

is the reaction-limited rate. k_R depends on the chemical reaction between the contacts at closest distance and is dominated by quenching. Therefore, we assume that k_R depends only on the temperature (T), while k_{D+} is dominated by diffusion and depends on the temperature as well as viscosity (η). By measuring k_{obs} for different η at particular temperatures, a linear relationship is obtained between $1/k_{obs}$ and η where k_R and k_{D+} can be determined from the y-intercept and the slope, respectively. Here k_{D+} will be in terms of η and generally normalized to the viscosity of water at that temperature.

2.2 SSS theory

Szabo, Schulten and Schulten (SSS) theory[44] derives the intra-molecular diffusion coefficient for the end-to-end distance of a chain, diffusing under the influence of a 1-D potential based on the Smoluchowski equation of diffusion. It considers diffusion between the end points as well as their reaction process at close contact. The potential takes the form $U(r) = -k_B T \ln P(r)$ where it depends on the distance probability distribution of the contacts, $P(r)$. Here k_B and T are the Boltzmann constant and temperature.

The theory defines the reaction-limited and the diffusion-limited rates as

$$k_R = \int_a^{l_c} q(r)P(r)dr = \sum_r q(r)P(r) \quad (Eq. 2.8)$$

and

$$\frac{1}{k_{D+}} = \frac{1}{k_R^2 D} \int_a^{l_c} \frac{dr}{P(r)} \left\{ \int_r^{l_c} (q(x) - k_R) P(x) dx \right\}^2 \quad (Eq. 2.9)$$

respectively, where a is the closest distance of approach of the contacts, l_c is the contour length of the chain, r is the distance between the contacts, m is the number of r values, $q(r)$ is the Trp-Cys contact quenching rate (**Eq. 2.1**) and D is the intra-molecular diffusion coefficient.

2.3 Pump-probe spectroscopy of Trp-Cys quenching

Pump-probe spectroscopy is a time-resolved spectroscopy technique that is used to measure ultrafast (nano to sub-microsecond) dynamics of a system. **Figure 2.2** shows the optical setup used here. A 266 nm UV pulsed laser beam (pump) with a 10ns excitation and a 10 ms relaxation time is created from the fourth harmonic of an Nd:YAG laser (Continuum Surelite II-10) and converted to 289 nm by a 1-m Raman converter (LIGHT AGE) filled with 450 PSI of D₂ gas. The sample is contained in a sealed long neck quartz cuvette (Hellma) with 10 mm path length and placed in a Peltier temperature-controlled sample holder (Quantum Northwest). The produced 289 nm beam is used to excite the Trp to a long-lived triplet state. Then the lifetime of the Trp population in the triplet state was probed between each pulse by transient absorption using a continuous wave

LASEVER 445 nm diode laser (probe). The pulsed beam is oriented at an angle of 10° from the probe beam to avoid detection by the detector and minimize background effects. The continuous laser is split into two prior to reaching the sample. One acts as the probe passing through the sample while the other acts as the reference. The two beam intensities are recorded simultaneously by two nanosecond photodetectors (New Focus) and amplified by a LeCroy DA1855A differential amplifier. The output is then fed into two digital oscilloscopes (Tektronix TDS 3032B) one recording the signal with millisecond resolution while the other records at microsecond resolution and the data is stored in a computer. Measurement and data collection process is controlled through a LabVIEW program.

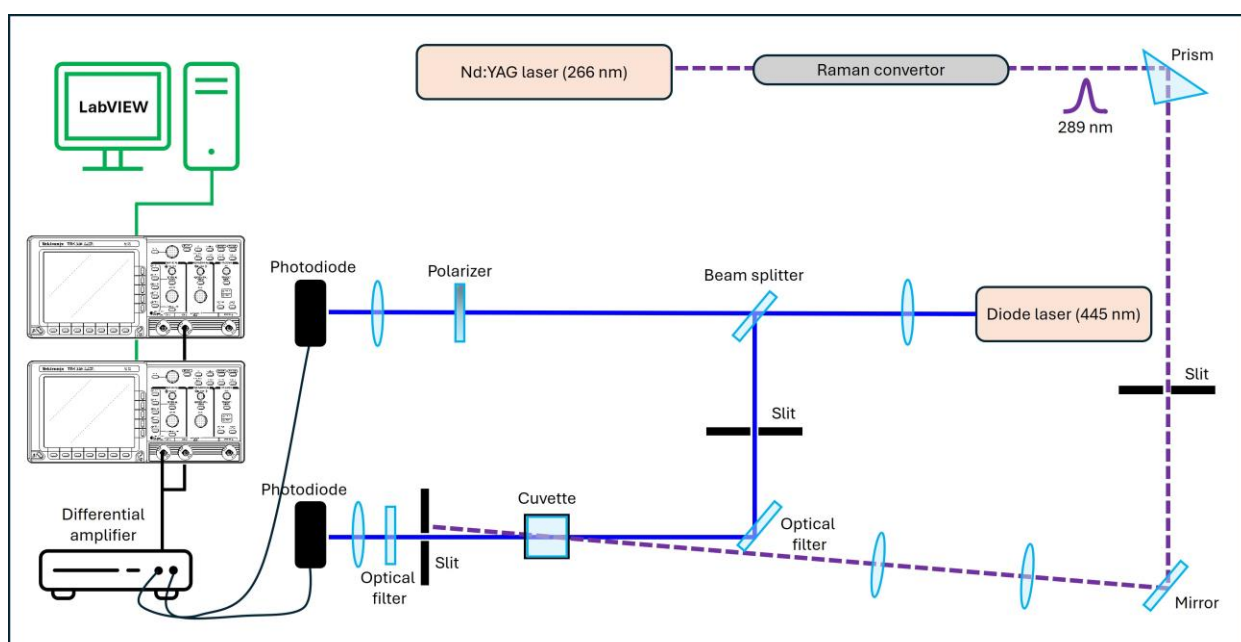


Figure 2.2: Pump-probe spectroscopy setup. Schematic shows the beam path of the 289 nm pulsed UV beam (dashed violet line) and the 445 nm continuous probe beam (solid blue line). The pulsed beam is set to cross the probe beam at an angle of 10° from the probe beam at the cuvette. The probe beam is kept at a power of 90 mW at the cuvette and the pulsed beam at a low power of ~ 5 mW to avoid photobleaching or degradation of the sample.

During measurements, first the instrument is aligned using a $50 \mu\text{M}$ NATA sample which is an uncharged analogue of tryptophan having a decay time of $\sim 20 \mu\text{s}$ at 20°C in 20mM sodium phosphate pH 8.0. After which the sample of interest is placed in the holder and the reference beam is adjusted using the polarizer to eliminate any DC voltage shifts. A background measurement is taken with just the pulsed beam to eliminate any pulse leakage and high frequency cable noise followed by the Trp-Cys decay measurement. The background measurement is then subtracted by the decay trace and saved in the computer.

Prior to measurements the sample is degassed using N_2O to remove any oxygen molecules and scavenge free electrons. We do this to eliminate any background contribution from photo effects such as the UV pulse generating hydrated electrons and neutral radicals that can absorb light near 450 nm and decay within 3 μs [35] and because O_2 is a good quencher of the tryptophan triplet state and can affect the decay measurements. And TCEP at $10 \times$ the concentration of the protein is added to reduce any disulfide bonds. Total sample volume used is 3 mL. At least 2.6 mL is degassed and the rest holds the protein which cannot be degassed due to the possibility of aggregation. The sample is deoxygenated (N_2O was let to flow on liquid surface) for ~ 10 minutes after protein addition. **Figure 2.3** shows the decay rate of Trp at different non-degassed volumes. Quenching remains stable from 300 – 450 μL indicating minimal background quenching. Above 450 μL quenching increases drastically probably due to either or both O_2 and solvated electrons. Therefore, we keep the non-degassed volume below 400 μL in all measurements.

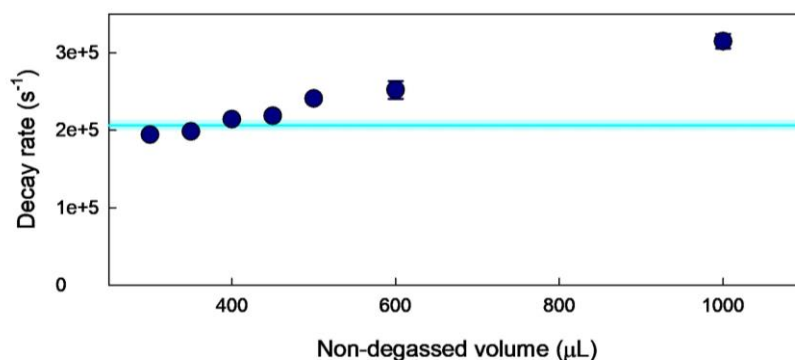


Figure 2.3: Background Trp quenching due to the non-degassed volume. Trp decay rates are shown for 30 μM α -syn39w69c at 10°C in 50 mM 13.33 mM NaCl Tris pH 7.4. A dramatic increase in the decay rate is observed above 450 μL due to Trp quenching by either or both oxygen and solvated electrons. From 300 μL - 450 μL the decay rate remains stable. Therefore, measurements are carried out below 450 μL of non-degassed volume to minimize this effect. The cyan belt indicates the average decay rate between 300 μL - 450 μL and its error.

2.4 Syringe pump system

A syringe pump system (**Figure 2.4**) was used to inject protein samples into the cuvette during pump-probe measurements (**Section 2.3**) as a method to minimize sample usage. Protein samples were injected into a 10 mm sealed quartz flow cuvette that can hold 800 μL of sample. Sample viscosity was controlled using sucrose, which was varied using three syringe-pumps (KDScientific-200). They were used to inject the protein, and deoxygenated buffers of 0 % and 50 % sucrose through a system of tubes where they mix before reaching the cuvette and further get

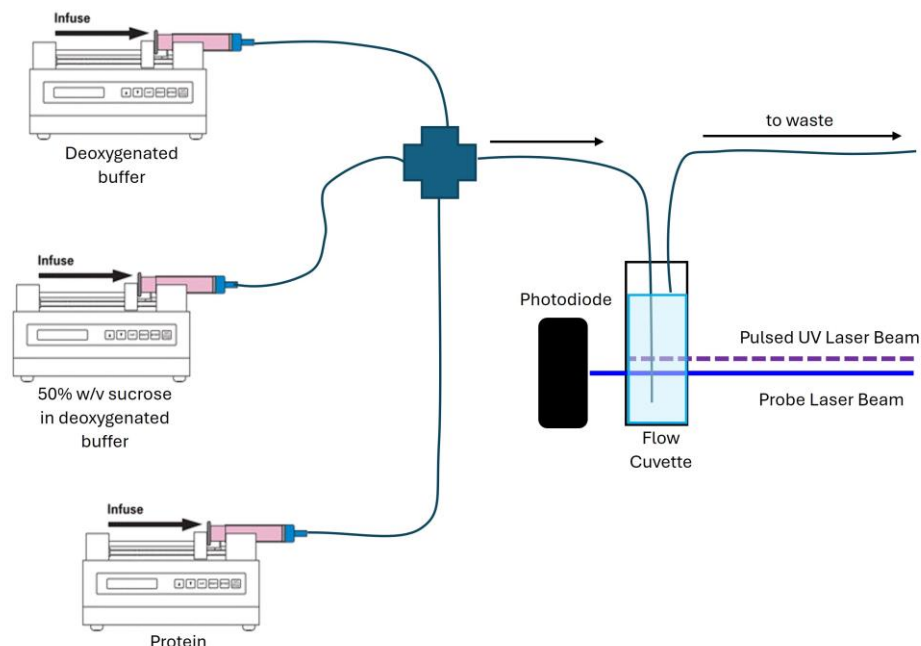


Figure 2.4: Pump-probe spectroscopy coupled with a syringe pump system. Syringe pumps are automated and controlled by a LabVIEW program where the protein and the buffers were injected into a sealed flow cuvette simultaneously.

mixed while inside the cuvette by a stir bar for 3 minutes before measurement. To prevent sample contact with air, the solutions are loaded into gas-tight syringes (Hamilton), and the cuvette is sealed. At the start of the measurements, air is flushed out by injecting 2 ml of buffer. During data collection, a total volume of 1.2 mL was injected, consisting of 0.2 mL 300 μ M protein, and a mixture of 0 % and 50 % sucrose solutions to create 0, 10, 20 and 30 % w/v of sucrose in the final sample. Every measurement was taken on a freshly injected sample and injected at a rate of 0.8 ml/min. Measurements were obtained within $\sim$ 1 hr of sample preparation and repeated twice. The viscosity of the injected volumes was confirmed using a BROOKFIELD DV-II+ Pro viscometer (Table 2.1).

Table 2.1: Measured viscosities of the samples for each sucrose percentage at 37°C.

Sucrose % w/v	Viscosity (cP)
0	0.68
10	0.90
20	1.20
30	1.70

2.5 Coarse-grained simulations

CG simulations were carried out using COCOMO (CONcentration-dependent CONdensation MOdel)[45] which is a residue-based CG model designed to describe phase separation in peptide only and peptide-RNA systems. COCOMO was developed by the Michael Feig group at MSU. Each residue is considered as a spherical bead with a volume equivalent to the residue volume. The model is parameterized to reproduce experimentally observed single-chain properties such as radius of gyration (R_g) as well as the concentration-dependent phase separation of a system[45].

Simulations were performed using OpenMM 7.7.0[46]. Langevin dynamics was applied with a friction coefficient of 0.01 ps^{-1} , with temperature fixed at 298 K. The energy of the systems was initially minimized using 5,000 steps of steepest descent followed by 20,000 steps of molecular dynamics with a time step of 0.01 ps. 5 minutes of clock time were spent to simulate 100 ns of a 100 residue protein.

The total interaction energy is given by:

$$\begin{aligned}
 U_{total} = & \sum_{i=1}^{N-1} \frac{1}{2} k_{bond} (l_{i,i+1} - l_0)^2 + \sum_{i=1}^{N-2} \frac{1}{2} k_{angle} (\theta_{i,i+1,i+2} - \theta_0)^2 \\
 & + \sum_{i,j} 4(\epsilon + \epsilon_{cation-\pi}) \left(\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{10} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^5 \right) \\
 & + \sum_{i,j} \frac{(A_{i,j} + A_{0i,j})}{r_{i,j}} e^{-\frac{r_{i,j}}{\kappa}}
 \end{aligned} \tag{Eq. 2.10}$$

where a bond potential accounting for chain connectivity, an angular potential between three neighboring residues accounting for chain stiffness, a short-range 10-5 Lennard-Jones potential and a long-range Debye-Hückel potential represents each term respectively. Optimized bonded and non-bonded interaction parameters are given in **Table 2.2** below. Non-bonded interactions were truncated at a cutoff distance of 3 nm and calculated using periodic boundary conditions. Residues separated by one bond were excluded from non-bonded interactions.

Table 2.2: Bonded and non-bonded interaction parameters used in COCOCMO and COCOMO2. Table A3 reports all residue-specific parameters used in simulations.

	Parameter	Description	Value	
			COCOMO	COCOMO2
Bonded interaction parameters	l	Distance between two neighboring residues		
	l_0	Equilibrium bond length	0.38 nm for peptides and 0.50 nm for RNA	
	k_{bond}	Spring constant	4184 kJ/mol·nm ²	
	θ	Angle between three beads		
	θ_0	Equilibrium angle	180°	
	k_{angle}	Angular constant	4.184 kJ/mol·rad ²	
Non-bonded interaction parameters	$r_{i,j}$	Inter-particle distance		
	$\sigma_{i,j} = 0.5(\sigma_i + \sigma_j)$	Distance at which the potential is zero. The effective radii $\sigma_i = 2r'_i \times 2^{-1/6}$, where r'_i is the radius of a sphere with equivalent volume of the residue i	Table A3	
	ε	Depth of the potential well	$\varepsilon_{polar} = 0.4 \text{ kJ/mol}$ $\varepsilon_{non-polar} = 0.4 \text{ kJ/mol}$ $\varepsilon_{nucleotides} = 0.4 \text{ kJ/mol}$	$\varepsilon_{polar} = 0.176 \text{ kJ/mol}$ $\varepsilon_{non-polar} = 0.295 \text{ kJ/mol}$
	$\varepsilon_{cation-\pi}$	Potential well depth adjustment for cation- π interactions	$\varepsilon_{R/K-F/Y/W} = 0.3 \text{ kJ/mol}$ $\varepsilon_{R/K-Neucleotide} = 0.2 \text{ kJ/mol}$	
	$A_{i,j} = A_i \times A_j$	Attractive or repulsive long-range interactions. $A_i = \text{sign}(q_i)\sqrt{0.75 q_i }$ [47] where q_i is the net charge of the residue i	Table A3	
	$A_{0i,j} = A_{0i} + A_{0j}$	Effective repulsion between residues due to solvation effects	$A_{0,polar/nucleotides} = 0.05$ $A_{0,non-polar} = 0$	$A_{0,polar} = 0$ $A_{0,non-polar} = 0.002$
	κ	Debye-Hückel screening length	1nm which corresponds to an ionic strength of ~100 mM	

The model uses a 10-5 Lennard-Jones potential over the 12-6 to make the potential ‘softer’, considering the larger particle sizes employed in CG simulations. A term for cation- π interactions is added which has been observed to contribute significantly to phase separation[48]. To account for the screening of residue-residues interactions by the ions in the solution, parameter A_0 was introduced in the long-range potential. And a screening length of $\kappa = 1$ nm was used which corresponded to ~ 100 mM of salt concentration. Hydrophobicity of the system is accounted for by using separate terms for polar and non-polar (hydrophobic) residues in short-range (ϵ) and long-range potentials (A_0).

An improved version of COCOMO, COCOMO2[2] was also used in this study. COCOMO2 extends the model’s applicability to phase separation of folded and multi-domain proteins. It is parameterized giving priority to saturation concentration of phase separation (C_{sat}) data from LLPS experiments in place of single-chain properties. The model shows a significant improvement in accurately predicting phase separation behavior, especially for longer IDPs compared to COCOMO. Since COCOMO2 is only used for IDPs in this study we only present the modifications relevant to IDPs. The form of the interaction potential remains same as indicated by **Eq. 2.10** while the short-range potential (ϵ) and the repulsion due to solvation effects (A_0) are re-parameterized as indicated in **Table 2.2**.

2.6 Expression of α -syn and RLP

Wild type α -syn plasmid was a kind gift from Gary Pielak (University of North Carolina, Chapel Hill, NC). The mutants α -syn39w69c, α -syn69c94w, α -syn-39w69c-v74e and α -syn-39w69c-a53t were created by mutagenesis[3]. The following method was used for the expression of all four proteins. Bacterial transformation was carried out using BL21(DE3) E. coli competent cells (Thermo Fisher Scientific) and α -syn recombinant plasmid DNA using the heat shock method. A culture tube each of 4 μ L α -syn plasmid, the positive control (4 μ L PUC19 DNA, Thermo Fisher Scientific) and negative control (blank) were incubated in ice for 30 minutes with 50 μ L of competent cells. Then the heat shock treatment was carried out by tapping the tubes in a 42°C water bath for 90 seconds and putting it back on ice for 2 minutes. After which the tubes were shaken for 1hr at 225 rpm in a 37°C incubation room. After shaking, 50 μ L and 1 μ L of α -syn, 100 μ L of positive and negative cultures were added separately to 4, 0.1 mg/mL ampicillin

(GOLDBIO) treated agar plates together with 50 μ L of S.O.C. medium. The liquids were spread on the plates using flame sterilized glass pipettes and incubated at 37°C overnight to grow.

8 tubes of 0.1 mg/ml ampicillin treated starter cultures were prepared with 5 ml of 25 g/L Luria Broth (Miller's LB, Invitrogen) in each tube. An E. coli colony was picked with a flame sterilized inoculating loop and dropped into each tube and shaken for 6 hrs at 225 rpm at 37°C in an incubation room. After shaking, 2 tubes each were inoculated into 4, 6 L flasks carrying 2 L of LB treated with 0.1 mg/ml ampicillin. The flasks were disinfected by autoclaving for 50 minutes at 121°C prior to the addition of ampicillin and starter cultures. The flasks were shaken in the 37°C incubation room for 15 hrs at 110 rpm. After shaking protein expression was induced by adding 2 ml of 1 M IPTG (GOLDBIO) to each flask and further shaken for 8 hrs. Next the cultures were centrifuged (SORVALL RC6 PLUS) at a speed of 380 g for 15 minutes at 4°C in Nalgene bottles (Thermo Fisher Scientific) and the pellets were collected and stored overnight at -20°C. Next the pellets were dissolved in 10 mL of Lysis buffer (50 mM sodium phosphate, 300 mM NaCl pH 8.0) in each Nalgene bottle, poured into a beaker and spatula tips of DNase (2000 Kunitz units/mg protein, Sigma), RNase (Roche), lysozyme (39000 units/mg protein, Sigma) and 2 tablets of protease inhibitor (cOmplete Tablets EDTA-free, Roche) were added. Then the cells were lysed by sonicating for 6 minutes using a Qsonica probe sonicator (Settings: Pulse 15 s 15 s Amplitude 35%) in ice while stirring. The lysed cells were transferred to centrifuge tubes and heated in a 100°C water bath for 30mins and centrifuged at a speed of 30310 g for 20 minutes at 4°C. After centrifuging, the supernatant was collected, and the protein was precipitated by the gradual addition of ammonium sulphate (Sigma) while stirring in ice. 361 mg of ammonium sulphate was added to every 1 mL of the supernatant. After the addition the mixture was further stirred in ice for 1hr for a better yield. Then the mixture was centrifuged at 34220 g for 25 minutes at 4°C, the supernatant was discarded, and the protein pellets were stored at -20°C overnight. Next, the pellets were desalted in 20 mM sodium phosphate buffer pH 8.0 by centrifuging at a speed of 6150 g for 25 minutes at 4°C using 3K Amicon ultra centrifugal filter units (Sigma). The process was repeated multiple times for optimum results.

After desalting the protein was purified by anion exchange chromatography followed by size exclusion chromatography using an ÄKTA FPLC (Fast protein liquid chromatography) system at

4°C. Anion exchange chromatography was carried out using a 5 mL HiTrap QFF column (Cytiva), with 20 mM sodium phosphate buffer pH 8.0 with a 1 M NaCl gradient. Size exclusion chromatography was carried out using a HiPrep 16/60 Sephacryl S-200 HR column with 20 mM sodium phosphate buffer pH 8.0. Protein was fractionated into tubes containing 20 μ L of 100 mM TCEP to reduce any disulfide bonds that can form due to the presence of cysteine. After purification the protein was concentrated using 2K centrifugal filter units (Vivaspin, SARTORIUS) into 600 μ M, 250 μ L aliquots.

The resin like polypeptide SKGP-(GRGDSPYS)₂₀-GY was expressed using recombinant plasmid DNA kindly donated by Ashutosh Chilkoti (Duke University, Durham, NC). Bacterial transformation was carried out using BL21(DE3) E. coli competent cells by the heat shock method similarly to α -syn expression as mentioned above except that the E. coli colonies were grown in 45 μ g/mL Kanamycin (GOLDBIO) resistive Agar plates.

5 ml starter cultures were prepared using 47 g/L of Terrific broth containing 45 μ g/mL of Kanamycin and incubated overnight for ~17 hrs at 37°C followed by inoculation into 1 L Terrific broth flasks containing 45 μ g/mL of Kanamycin and incubated in a shaker (110 rpm) at 37°C for 9 hrs. Protein expression was induced with 500 μ L of 1 M IPTG and shaken for 18 hrs at 37°C. The cultures were centrifuged at 3,500 g for 15 minutes at 4°C and the pellet was resuspended in 10 ml milli-Q water. Cells were lysed by sonication (2 minutes, 10 s pulsing, 40 s rest) and centrifuged at 20,000 g for 20 minutes at 4°C. The pellets were suspended in 10 ml of 4 M Urea + 150 mM PBS buffer pH 7.4 and heated for 10 minutes in a 37°C water bath. The suspension was centrifuge at 20,000 g for 20 minutes at 4°C and the supernatant was dialyzed in a 10 kDa membrane (Thermo Fisher Scientific SnakeSkin Dialysis Tubing) against 1:200 milli-Q water at 4°C for 48 hrs. After dialysis, the suspension in the bag was collected and centrifuged at 3,500 g for 10 minutes at 4°C. Then the pellet was lyophilized for 24 hrs.

2.7 Fluorescent labeling of α -syn and RLP

α -syn^{39w69c} was fluorescently labeled with Alexa Fluor 488 C₅ Maleimide (Thermo Fisher Scientific). The dye is attached to the α -syn cysteine via maleimide conjugation forming a stable thioether bond. Purchased Alexa-488 maleimide was dissolved in anhydrous DMSO (Invitrogen)

to a concentration of 10 mM to prepare a stock solution. Expressed α -syn^{39w69c} (**Section 2.6**) was buffer exchanged into 20 mM sodium phosphate pH 7.4 and brought to 300 μ M using Nanosep 3K Omega (PALL) centrifugal filter units. TCEP 10 $\times$ the concentration of the protein was added and let to rest for 20 minutes to reduce disulfide bonds. Then the protein was mixed with Alexa-488 maleimide at a molar ratio of 1:15 and incubated overnight at 4°C protected from light for the reaction to take place. After incubation the sample was diluted to 500 μ L in 20 mM sodium phosphate pH 7.4 and purified by FPLC using 3 Sephadex G-25 size exclusion columns (Cytiva HiTrap Desalting, 5 mL) in series with 20 mM sodium phosphate pH 7.4. After the run the labeled protein was recovered from the early fractions while the free dye eluted later.

RLP was fluorescently labeled with Alexa Fluor 647 NHS ester (Thermo Fisher Scientific) which attaches to the primary amine of the protein. Purchased Alexa-647 NHS ester was dissolved in anhydrous DMSO (Invitrogen) to a concentration of 15mM to prepare a stock solution. A 300 μ M RLP sample was prepared in 20 mM sodium phosphate pH 7.4 as indicated in **Section 2.6** and diluted down to 100 μ M. The sample is then mixed with Alexa-647 NHS ester at a molar ratio of 1:15. The mixture was stirred gently using a magnetic stir bar for 1 hr at room temperature protected from light for the reaction to take place. After conjugation the sample was brought to 500 μ L by adding 8 M urea in 20 mM sodium phosphate pH 7.4 and size exclusion chromatography was carried out using the 8 M urea with 3 Sephadex G-25 columns in series. The labeled protein fraction was separated, and dialysis was carried out using a 10 kDa Snakeskin membrane against 1:200 20 mM sodium phosphate pH 7.4 at 4°C for 48 hrs. After dialysis, the sample in the bag was collected.

2.8 Dynamic light scattering of condensates

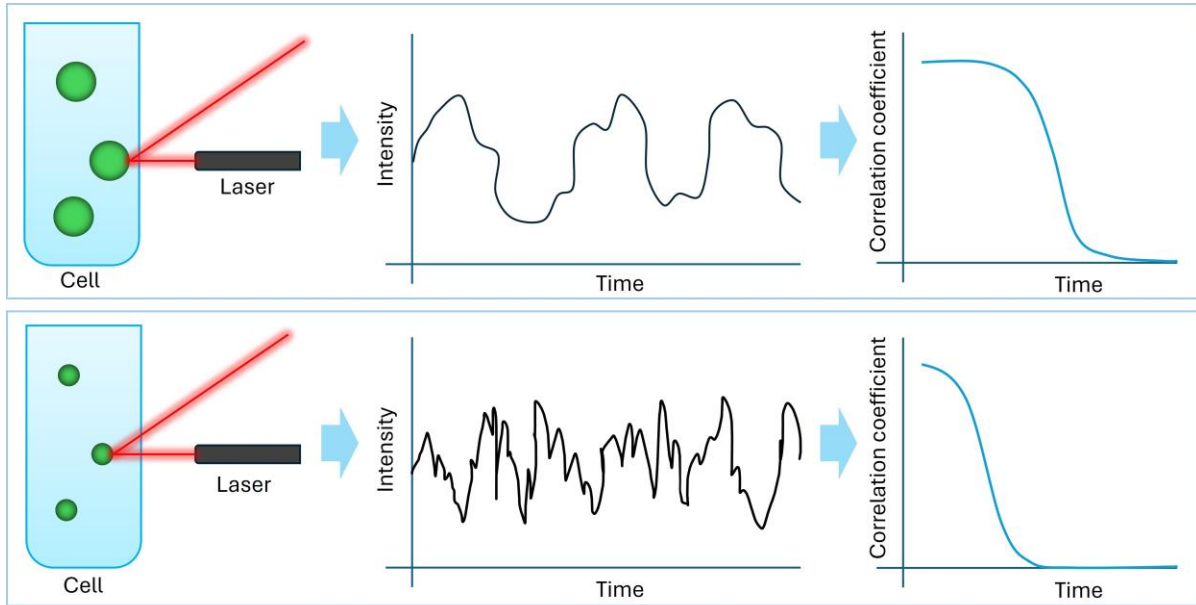


Figure 2.5: Schematic outlining the DLS technique. The upper panel indicates the measurement of larger particles resulting in slower intensity fluctuations and a longer correlation. The lower panel indicates the measurement of smaller particles resulting in faster intensity fluctuations and a rapid correlation.

Dynamic light scattering (DLS) is a technique used to determine the size distribution of particles or molecules. It uses the Brownian motion of the particles to detect their hydrodynamic size. Particles are illuminated by a laser and the scattered light is measured at different angles. The intensity of the scattered light will fluctuate over time with the diffusion of the particles due to Brownian motion. The smaller the particles, faster the diffusion, leading to rapid intensity fluctuations and the larger the particles, slower the diffusion, leading to long intensity fluctuations. Fluctuations are recorded over time forming an intensity autocorrelation function. Larger particles give a longer correlation, and smaller particles give a rapid correlation.

The autocorrelation function is fitted to an exponential decay which gives the decay rate Γ . Using the equation below the diffusion coefficient D is calculated.

$$\Gamma = q^2 D \quad (\text{Eq. 2.11})$$

where q is the wave vector.

$$q = \frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad (\text{Eq. 2.12})$$

where λ is the wavelength of the laser, n is the solvent refractive index and θ is the scattering angle. The hydrodynamic diameter is calculated using the Stokes-Einstein equation.

$$d_H = \frac{kT}{3\pi\eta D} \quad (\text{Eq. 2.13})$$

where T is the temperature and η is the solvent viscosity.

2.9 Isothermal titration calorimetry of condensate formation

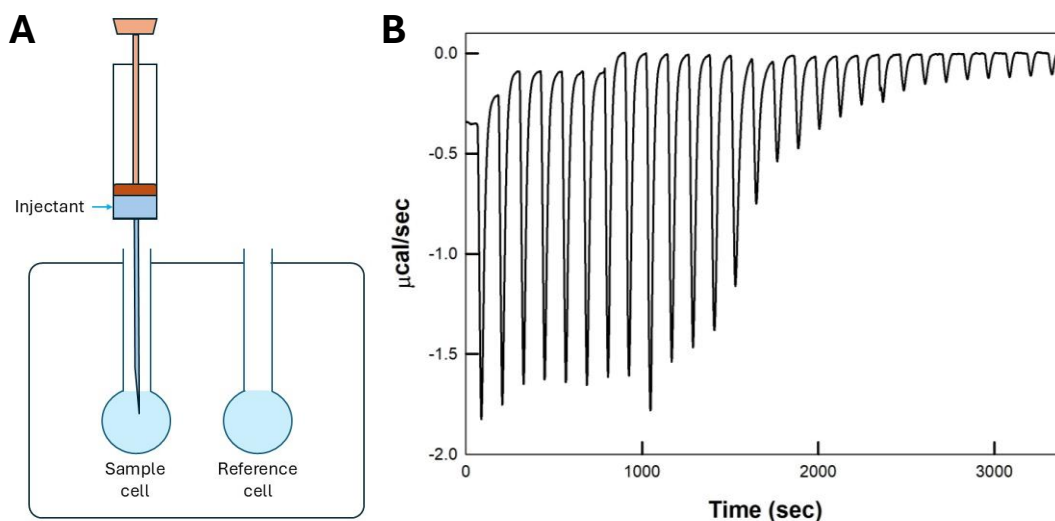


Figure 2.6: ITC setup and recorded heat measurements. **A)** The diagram shows the ITC setup where the injectant is titrated into the solution in the sample cell. Measurements are taken while keeping the sample and the reference cell at the same isothermal conditions where the reference cell is loaded with the sample buffer. **B)** Rate of heat released per injection over time for PolyA₂₀ and [RGRGG]₁₀ mixture. PolyA₂₀ was placed in the cell and [RGRGG]₁₀ was injected in volumes of 10 μL .

ITC is a technique used to measure the enthalpy of a reaction between two different types of molecules. One is placed in a cell while the other is injected into the cell in small quantities repeatedly (**Figure 2.6.A**). The instrument detects the heat absorbed or released as positive and negative values respectively during molecular interaction for each injection and the rate of heat change is recorded with time as shown in **Figure 2.6.B**. The peaks diminish gradually with time as the molecules in the cell get saturated by the injectant and only the heats of dilution remain.

Both molecules should be in the same buffer to avoid any heat changes due to ionizing effects that can arise when mixing. The recorded data is then baseline subtracted, and the peaks are integrated to obtain the binding isotherms.

Chapter 3

Length dependence of liquid-liquid phase separation

In this chapter we investigate the length dependence of LLPS and its underlying thermodynamics using a short polymer system of various lengths of the RNA polymer PolyA with repeats of the peptide RGRGG. Condensates are visualized through confocal microscopy. CG simulations and various experimental techniques are used to analyze the thermodynamics of the system.

3.1 Phase separation of the peptide-RNA system

COCOMO CG molecular dynamic simulations (**Section 2.5**) were carried out for different lengths of PolyA_N (N = 5, 10, 20, 300) with various repeats of [RGRGG]_M (M = 1, 2, 3, 4, 6, 8, 10) by the Feig group. The peptide and RNA were chosen to have opposite charges to facilitate LLPS. Here polyadenine (PolyA) has a negative charge on every base, while RGRGG has a positive charge on every arginine (R). An initial random conformation for each RNA or peptide chain was obtained using a custom python script and placed in a 100 nm size simulation box at random positions and with random orientations but avoiding any two residues between different molecules to be closer than 5 nm, until the required concentration is fulfilled. In all these mixtures the concentration of the peptide and the RNA was maintained at 1 mg/mL so that the total number of nucleic and amino acids stay constant. This made it possible for us to vary only the polymer length while keeping everything else the same. For example, comparing the mixtures formed by PolyA₅ with [RGRGG]₂ and [RGRGG]₁₀ there are 5 [RGRGG]₂s for each [RGRGG]₁₀ but the total number of amino acids remains the same.

The systems were simulated for 20 μ s with a time step of 0.02 ps under periodic boundary conditions with five replicates each. Coordinates were saved every 500 ps. We observed that shorter lengths are more reluctant to form condensates compared to the longer peptides and RNAs creating a phase boundary for LLPS depending on polymer length (**Figure 3.1**). For either polymer, longer lengths of one species were required to phase-separate shorter lengths of the other. These simulation results are in general agreement with known behavior of LLPS of peptides and RNA where longer lengths facilitate LLPS ([49-51]).

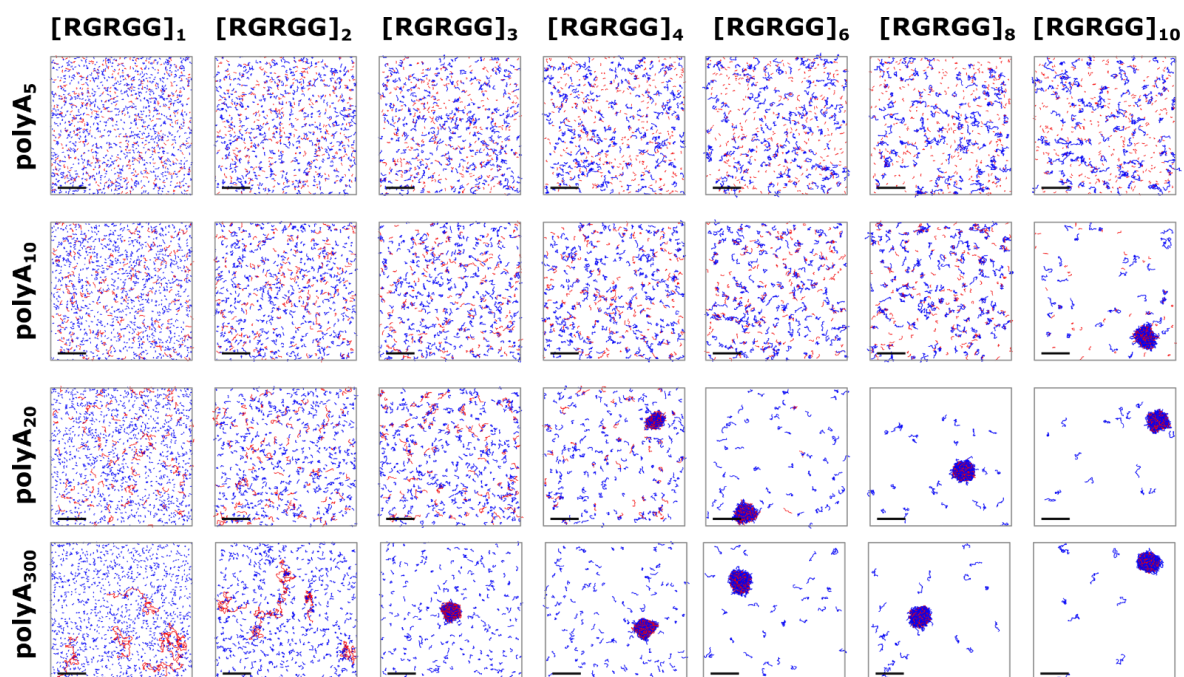


Figure 3.1: Length-dependent LLPS of different protein-RNA mixtures using COCOMO. Simulations were performed using COCOMO CG model. For each of these mixtures the final frame of the trajectory is shown here where RNA and peptides are colored in red and blue, respectively. Scale bars represent 20 nm.

To confirm these predictions, confocal microscopy measurements were carried out for mixtures of PolyA_N (N = 5, 10, 20, > 600) and [RGRGG]_M (M = 1, 2, 3, 4, 6, 8, 10) (**Figure 3.2**). Lyophilized powders of the peptides [RGRGG]_{1,2,3,4,6,8,10} and Cy5 labeled [RGRGG]₁ were purchased from Bio-Synthesis and the RNAs PolyA_{5,10,20} and Cy3 labeled PolyA_{10,20} from Horizon Discovery. Powders were initially dissolved in 20 mM Sodium Phosphate buffer pH 8.0 to 1 mM or 5 mM in case of unlabeled and 400 μ M in the case of labeled peptide and RNA and used as stock solutions from which the mixtures were prepared maintaining the concentrations of all species at 1 mg/mL and the positive to negative charge ratio of each mixture at 0.75 unless and otherwise mentioned. The samples were prepared by simply adding together the required amounts of peptide and the RNA from the stock solutions and pipette mixing. Cy3-labeled PolyA was used as the fluorophore at a low concentration of 5 μ M to avoid inducing phase separation (**Section A1**).

Confocal images were captured using a Nikon A1Rsi Confocal Laser Scanning Microscope configured on an automated Nikon Eclipse Ti inverted microscope equipped with a 100 $\times$ Plan Apo total internal reflection fluorescence (TIRF) oil objective (NA 1.45) at 100 $\times$ objective magnification and photomultiplier tube (PMT) detector set to 31 HV. The Cy3 and the Cy5 were

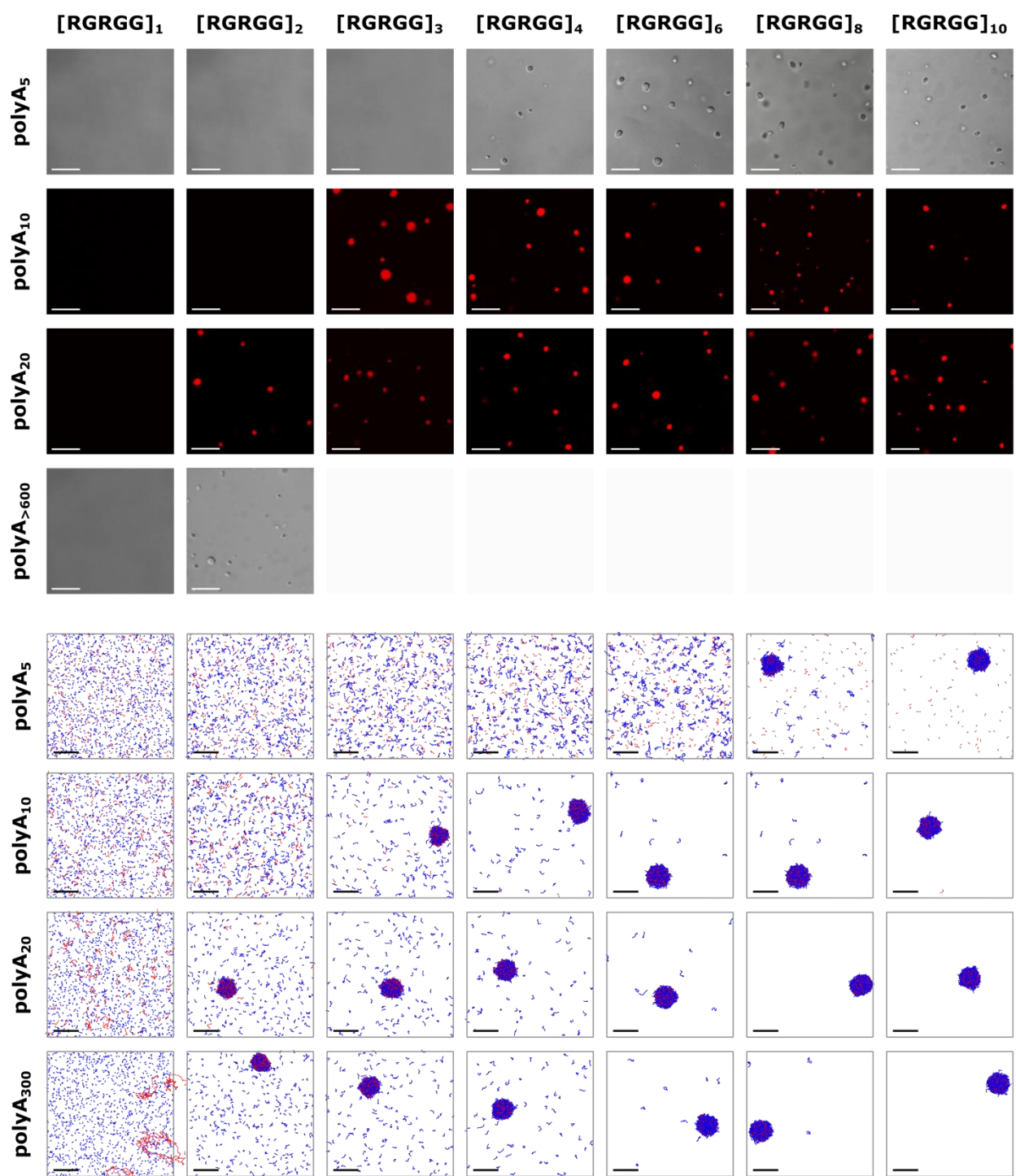


Figure 3.2: Experimental and COCOMO 1.2 σ results for LLPS of different peptide-RNA mixtures. Upper panels show experimental results of the protein-RNA mixtures obtained by confocal and differential interference contrast microscopy. 5 μ M Cy3-labeled PolyA was used for fluorescence. The lower panels show the simulation results using COCOMO 1.2 σ . RNA and protein are colored in red and blue, respectively. The polymers are at 1 mg/mL in both experiment and simulation. Scale bars represent 10 μ m in experimental panels and 20 nm in the simulation panels.

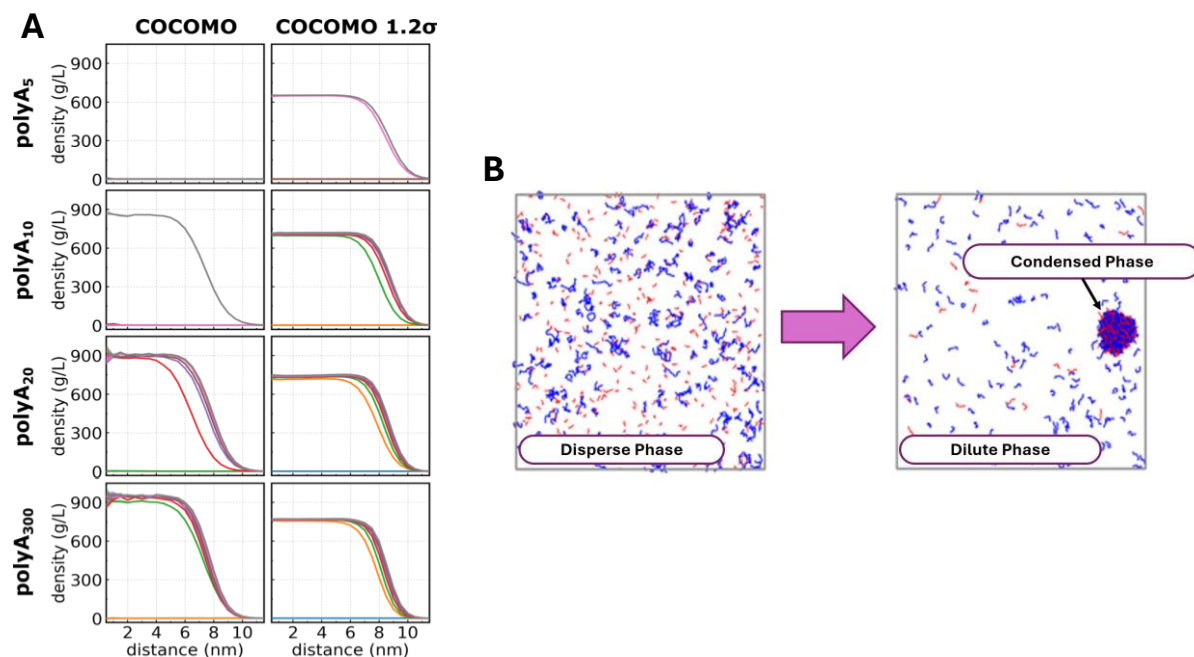


Figure 3.3: A) Cluster density in condensates. The collective density of peptide and RNA is shown as a function of distance, from the center of the condensate outwards. The left column shows the high densities in COCOMO simulations, and the right column shows the reduced densities achieved in COCOMO 1.2σ by increasing the size of the beads by ~ 20 %. Each row represents different RNA lengths while different colored traces represent peptide lengths. Blue, orange, green, red, purple, brown, and pink, represent $M = 1, 2, 3, 4, 5, 6, 8$, and 10 for $[RGRGG]_M$ respectively. Flat density lines indicate that no clusters were formed in that particular system. Initial concentrations for these simulations were kept at 1mg/mL for both protein and RNA. **B) Phases.** The diagram shows the different type of phases encountered in this study.

excited using a diode laser at 561 nm and 647 nm and fluorescence was detected through 595/50 nm and 700/75 nm band-pass emission filters respectively. Transmitted light images were recorded using Differential Interference Contrast (DIC) optics at 561 nm. Image acquisition was performed using the Nikon NIS Elements software (version 5.21.03). Results were mostly in agreement with simulation data except for the shift in the phase boundary (**Figure 3.2**). As predicted by simulations, longer lengths of either the peptide or RNA were needed to phase-separate the shorter of the other confirming the need for a minimum polymer length for phase separation. This behavior is well highlighted with the shortest peptide $[RGRGG]_1$ which doesn't seem to phase-separate at all, even with the extremely long chain PolyA_{>600}. This indicates that the peptide is too short to sufficiently interact to form stable condensates.

The simulations were further analyzed to understand the discrepancy in the phase boundary. It was noticed that the density in the condensed phase was too high (~900 g/L), which implies there is

little water inside the condensate (**Figure 3.3.A**). This is not the typical behavior of condensates, where experimentally we observe condensate growth over time indicating its liquid nature (**Section 3.4**). Increasing the size of beads by $\sim 20\%$ we were able to observe a significant decrease in the density (~ 750 g/L) retaining the liquid nature of the condensates. At the same time this shifted the phase boundary of the simulation data to lower lengths in quantitative agreement with experiment. This modification of bead size increases the effective radii, σ_i (**Table 2.2**) by a factor of 1.2 hence we redefine the model as COCOMO1.2 σ and is used for the rest of the simulations in this chapter. The redefined residue-specific σ_i are listed in **Table A3**.

For convenience the phase after condensation, phase before condensation and the phase coexisting with the condensed phase are known here as condensed phase, dispersed phase and the dilute phase respectively (**Figure 3.3.B**). All systems were simulated under periodic boundary conditions. To confirm the observed phase separation is not affected by the periodicity of the systems, simulations were repeated for several PolyA_N-[RGRGG]_M systems with a larger box of 200 nm using COCOMO1.2 σ (**Figure A5**). There were no changes in the results, and the phase boundary in **Figure 3.2** was retained as well. We did observe that the condensates were larger in size. This is possible due to the presence of a larger number of molecules.

3.2 The phase boundary

We clearly observe a sharp phase boundary that depends on peptide and RNA lengths (**Figure 3.2**). To be certain about the observed phase boundary and the length dependent behavior, particle size analysis was carried out using dynamic light scattering (**Section 2.8**) to check the existence of any condensates which the confocal may not have captured due to its diffraction limit of the confocal microscope (~ 400 nm) using a Zetasizer Ultra Red Advanced Series instrument. Samples were contained in a Hellma 45 μ L quartz cuvette of 3mm path length. The measurements were taken at 25°C using the 173° backscatter detector at a wavelength of 633 nm. Particle size is reported in terms of the hydrodynamic diameter calculated using the Stokes-Einstein equation (**Eq. 2.13**) for a refractive index of 1.33 and a viscosity of 0.89 mPa of water. The backscattering function was used to reduce the photon path length in the sample. This was done to avoid multiple scattering events that will arise due to the high turbidity and the high particle concentrations.

Less than 10 nm size particles were observed (**Figure 3.4**) for mixtures that don't show any condensates in confocal microscopy which is the monomer size range for the peptides/RNAs used. Large particle sizes > 700 nm were observed for mixtures that show condensates in confocal microscopy. These results are consistent with confocal measurements with one exception of PolyA₅[RGRGG]₃ consisting of particles ~ 150 nm while no condensates were observed with DIC. These may be unstable condensates forming at the phase boundary which are unable to grow within the time limit of the confocal measurements. Therefore, the phase boundary observed in confocal microscopy is generally correct.

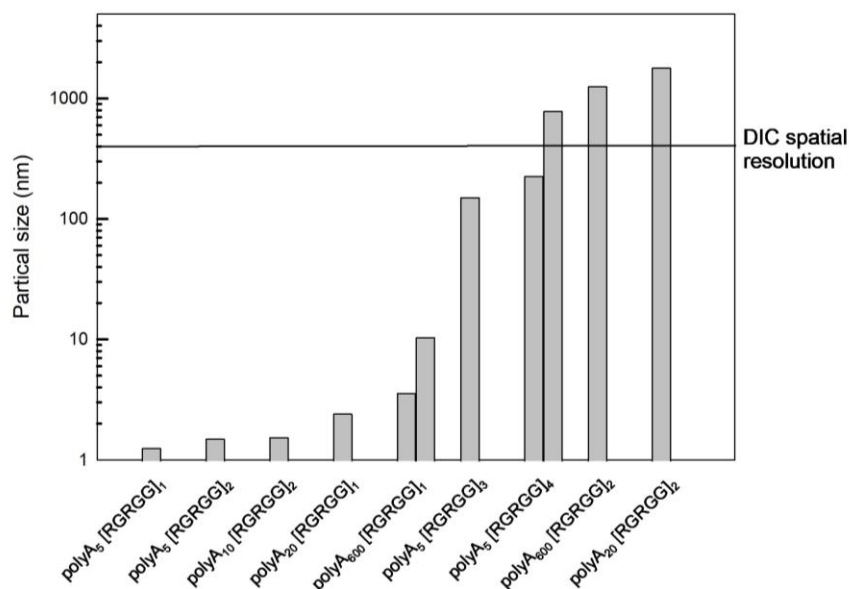


Figure 3.4: Particle size of protein-RNA mixtures measured by DLS. Small particles < 10 nm were detected below the DIC diffraction limit of the confocal while larger particles > 700 nm were detected above showing a large size separation with the one exception of PolyA₅[RGRGG]₃ which was ~ 150 nm in size. Here the particle size is reported in terms of mean hydrodynamic diameter obtained from particle size distributions (**Figure A6**).

3.3 Condensate growth

For all length combinations of peptide and RNA the confocal images were typically captured within 3 minutes of sample preparation. To study the growth of condensates with time 10 images were captured for several of these length combinations within ~ 2 minutes, right after sample preparation and then ~ 10 minutes and ~ 20 minutes after sample preparation. All three, time measurements were taken using the same sample and the aliquot was well mixed with the pipette before each measurement to ensure that droplets are resuspended. Each of these 10 images were captured in 10 separate areas of a drop (20 μ L) of sample. The size distribution analysis of

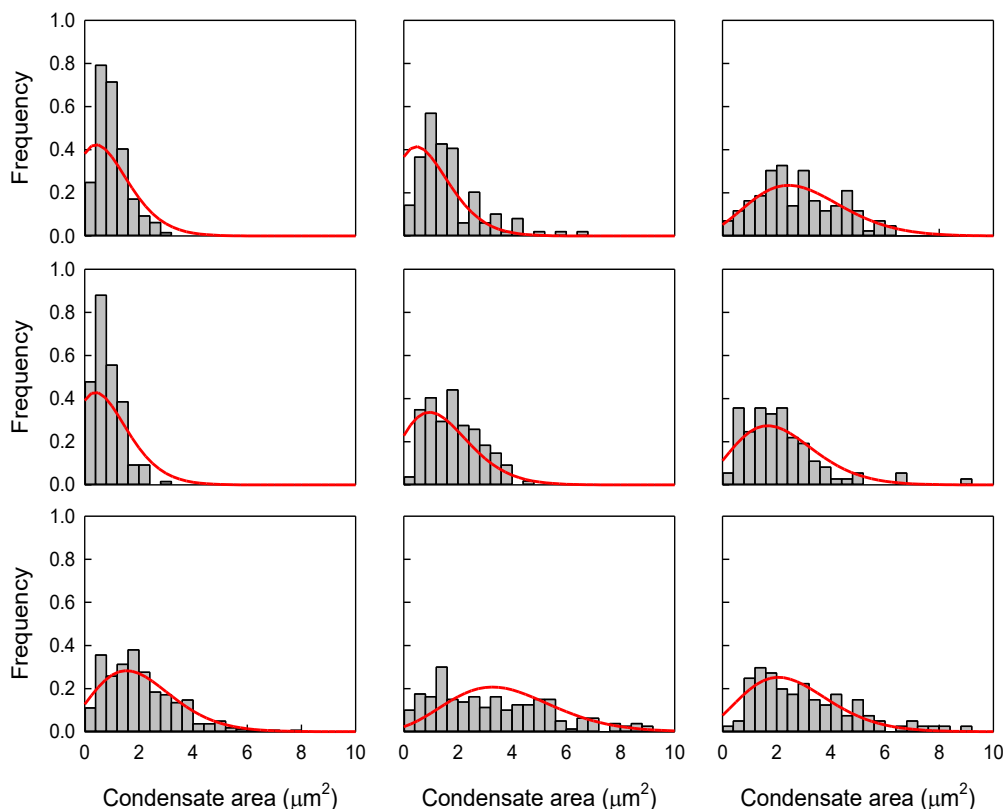


Figure 3.5: Size distribution of condensates over time. Top row: PolyA₁₀ and [RGRGG]₄ around (from left) 4:15, 11:05 and 20:56 minutes after mixing. Middle row: PolyA₁₀ and [RGRGG]₁₀ around 3:23, 10:50 and 20:44 minutes after mixing. Bottom row: PolyA₂₀ and [RGRGG]₄ around 3:32, 11:54 and 21:11 minutes after mixing. The bin width is 0.4 μm^2 . The red lines fit a Poisson distribution. Concentrations are the same as in **Figure 3.2**.

condensates was performed using the ImageJ software. The area of each fully visible condensate was calculated. Partially visible condensates at the edge of the images were discarded. Condensates were binned according to the area at a bin width of 0.4 μm^2 .

One might expect longer chains to have a higher probability of interacting due to the presence of more binding sites compared to shorter chains resulting in an increase of droplet size with length. But we do not see an increase in droplet size with length in this system. Droplet size was measured for several different mixtures of PolyA_N and [RGRGG]_M using confocal to observe their growth. Somewhat broader distribution of sizes was observed for all samples ranging from ~ 0.5 to ~ 5 μm . The histograms were fitted with Poisson distributions. A clear shift of the average to higher droplet sizes over time shows that the condensates seem to grow, and almost to twice their size within a short period of ~ 20 min. But we do not see a change in droplet size with polymer length within the uncertainty of the distribution. This suggests that the residue composition of

condensates remains the same for different polymer lengths. Each peptide-RNA system can have the same number of interactions when forming the condensates, and the enthalpy of phase separation can be the same from one mixture to another.

3.4 Liquid nature of the condensates

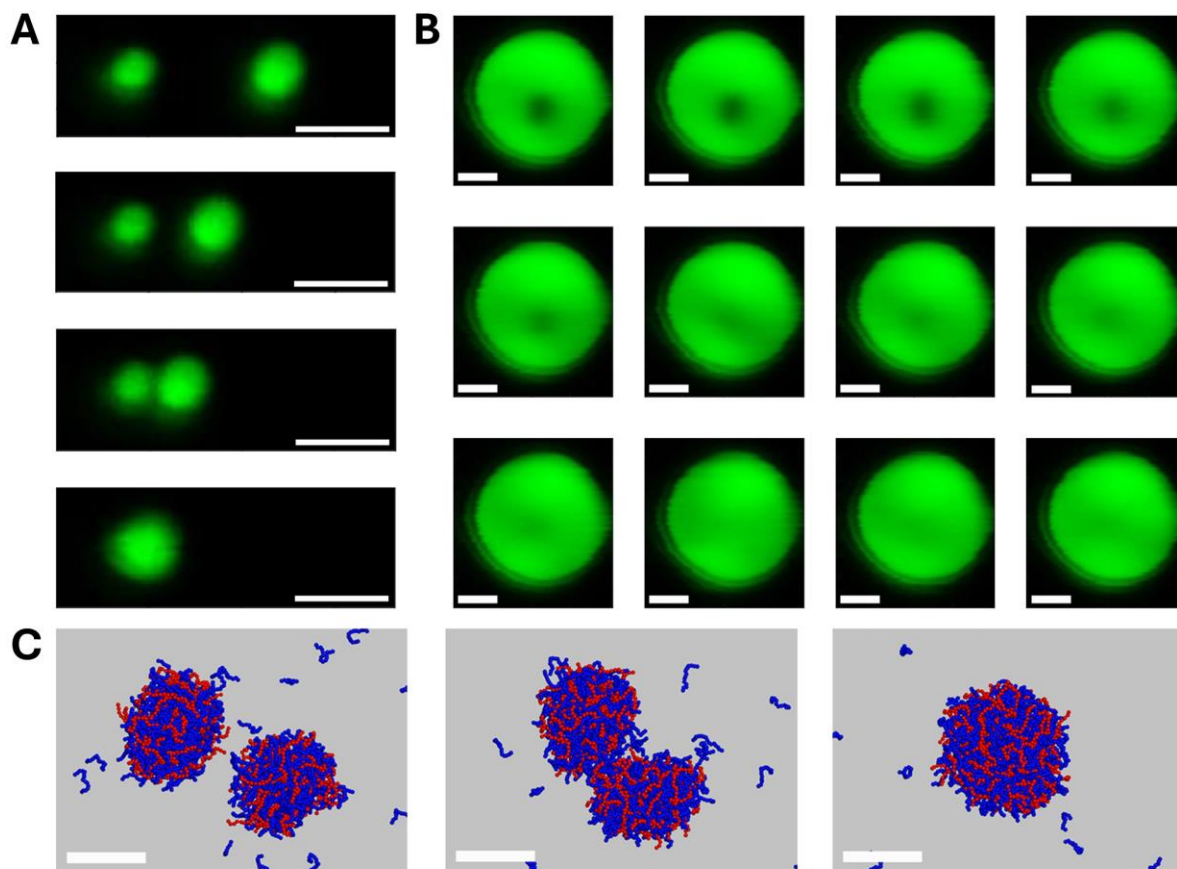


Figure 3.6: Liquid behavior of peptide-RNA condensates: **A)** Fusion of two PolyA₂₀-[RGRGG]₁₀ droplets. The two droplets are moved towards each other until fusion (From top to bottom) Scale bars: 2 μ m. **B)** FRAP showing initial photobleaching of a PolyA₂₀-[RGRGG]₁₀ condensate followed by 100 % recovery. Scale bars: 2 μ m. **C)** Fusion of two condensates in CG simulations. Scale bars: 10 nm.

Fusion and Fluorescence recovery after photobleaching (FRAP) measurements were carried out on PolyA₂₀-[RGRGG]₁₀ condensates at 1mg/ml using LUMICKS C-Trap (**Figure 3.6.A and B**). Measurements were carried out by LUMICKS USA Inc. 5 μ M of Cy3-labeled PolyA₂₀ was used as the fluorophore. To test fusion, two condensates were optically trapped and one trap moved towards the other and held at close proximity until they came into contact on their own due to subtle movements. Rapid fusion was observed. FRAP was performed by photobleaching a condensate for 3 - 5 s at 100% of laser power and observing the fluorescent recovery at a laser

power of 10 % where full recovery was observed. This experimental evidence together with droplet growth observed as shown in **Section 3.3**, confirms the liquid nature of the condensates.

Figure 3.6.C shows a fusion event between two PolyA₂₀-[RGRGG]₁₀ condensates in simulation indicating the liquid-nature of the studied systems. We also measured the residence time of polymers in the condensed phase which were less than 1/2 of the run time for the shorter systems and increased commensurately with the polymer length indicating the ease of movement between phases.

3.5 Induction of LLPS by short peptides

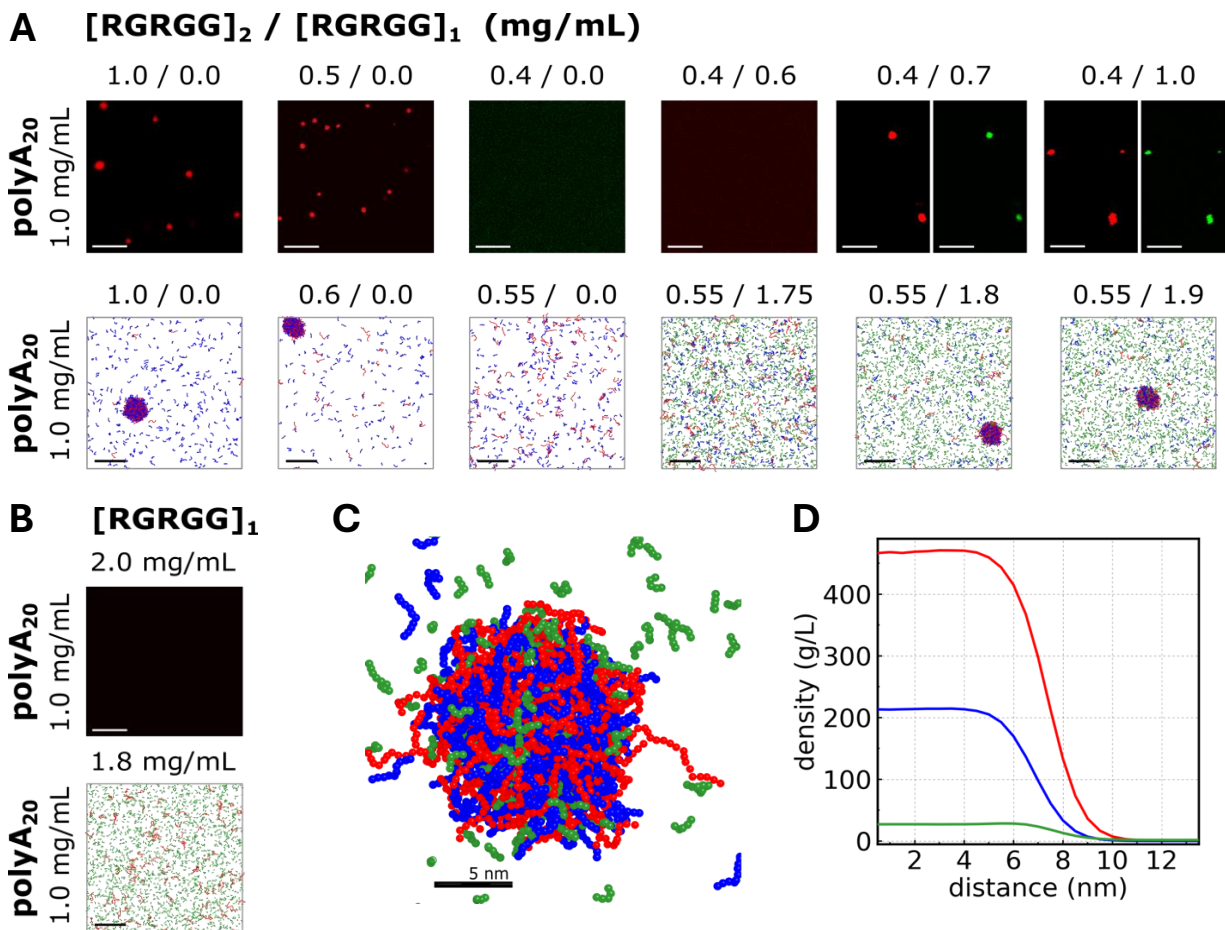


Figure 3.7: Phase separation recovery by short peptides. Here PolyA₂₀, [RGRGG]₁, and [RGRGG]₂ is indicated in red, green, and blue, respectively. PolyA₂₀ was kept at 1 mg/mL for all measurements. Scale bars represent 10 μ m in experimental panels (upper panels in A and B) and 20 nm in the simulation panels (lower panels in A and B). **A)** Experimental and simulation using COCOMO 1.2 σ results show the loss of LLPS when lowering [RGRGG]₂ concentration below a certain threshold and its recovery when adding sufficient [RGRGG]₁. In the upper panel [RGRGG]₂-Cy3 fluorescence is represented in red while the [RGRGG]₁-Cy5 fluorescence is represented in green indicating coexistence in the condensates. The lower panels show the final frames of the trajectories for each simulated system. **B)** Shows that [RGRGG]₁ doesn't induce phase-separate even at high concentrations which is supported by simulations. **C) and D)** Formed condensate in simulations and its density profile as a function of distance from the center of the condensate. All polymers are highly dense inside the condensate compared to the dilute phase. Initial concentrations of PolyA₂₀, [RGRGG]₂, and [RGRGG]₁ in this simulation were 1.0, 0.55, and 1.8 mg/mL, respectively.

Experimental and simulation data indicate minimum peptide and RNA lengths for phase separation. However, we see that the shortest peptide [RGRGG]₁ participating in condensation when the condensates are formed by longer peptides. This led us to speculate that shorter polymers may be able to compensate when the concentration of a longer polymer was too low to trigger condensation. To demonstrate this, we reduced the concentration of [RGRGG]₂ in a mixture with

PolyA₂₀ until no condensates were observed, which was achieved between 0.4 and 0.5 mg/mL. Then [RGRGG]₁ was added until phase separation was recovered. Recovery was observed between 0.6 and 0.7 mg/mL confirming the speculation (**Figure 3.7.A**). **Figure 3.7.B** shows that [RGRGG]₁ cannot induce phase separation by its own even at high concentrations of 2 mg/ml. These observations were confirmed by COCOMO 1.2 σ simulations where the threshold for [RGRGG]₂ condensation was observed at 0.55 mg/mL and phase separation was recovered at 1.8 mg/mL of [RGRGG]₁.

3.6 Enthalpy of phase separation

To understand the role of enthalpy in phase separation we used isothermal titration calorimetry (ITC). Measurements were carried out using a Micro-Cal VP-ITC system at 25°C for several of the mixtures including the ones near the phase boundary. 1.51 μ L of PolyA₁₀ and 2.75 μ L of PolyA₂₀ at $\sim$ 0.1 mg/mL were titrated with 30, 10 μ L, 250 μ M injections of [RGRGG]₄ and [RGRGG]_{1,4,8,10} respectively (**Table A1**). Samples were deoxygenated for a few minutes prior to measurements. Results show that the mixtures that phase-separate undergo a rapid enthalpy transition while the ones that do not, show no change in heat (**Figure 3.8**). Bright-field microscopy images captured for 0.1 mg/mL mixtures of RNA and peptide confirm the correlation between

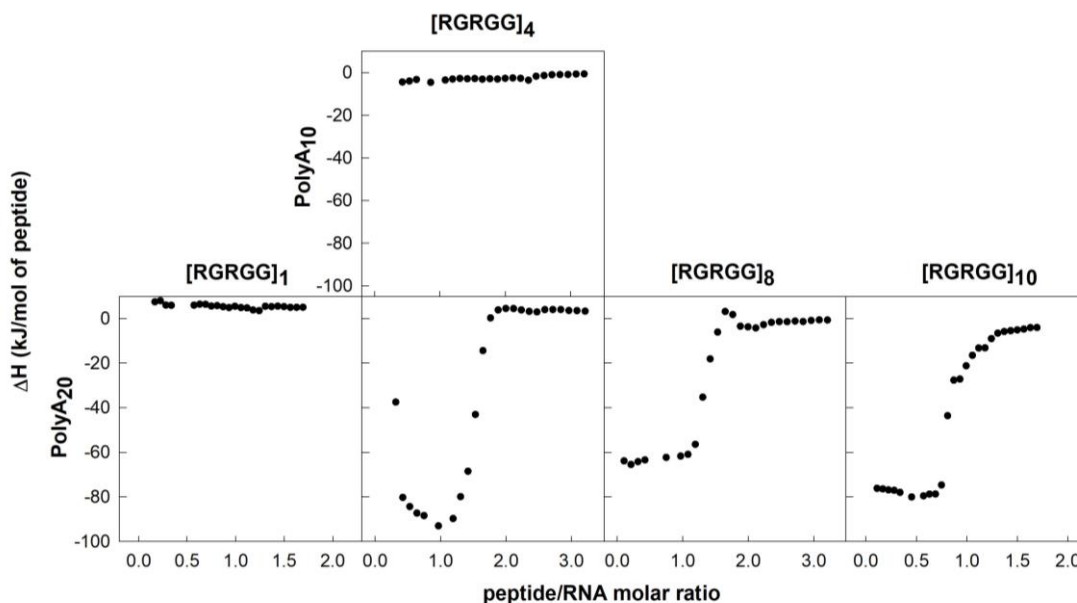


Figure 3.8: ITC binding isotherms for various mixtures of RNA and peptide. PolyA₂₀ – [RGRGG]₄, PolyA₂₀ – [RGRGG]₈ and PolyA₂₀ – [RGRGG]₁₀ shows a distinct enthalpy transition while PolyA₂₀ – [RGRGG]₁ and PolyA₁₀ – [RGRGG]₄ does not in agreement with the phase separation behavior of the mixtures where the former samples condense while the latter does not.

phase separation and the observed enthalpy transition (**Figure 3.9, Table A2**). Here we estimated the enthalpy of phase separation as the change in heat released before and after condensation, yielding enthalpies around -60 to -90 kJ/mol of peptide for all analyzed mixtures. A slight kink is observed near the transition which can be due to phase separation behavior[52]. However, for a more detailed thermodynamic analysis of the ITC curves, a specific binding model that applies well to the condensation of disordered polymers will be required.

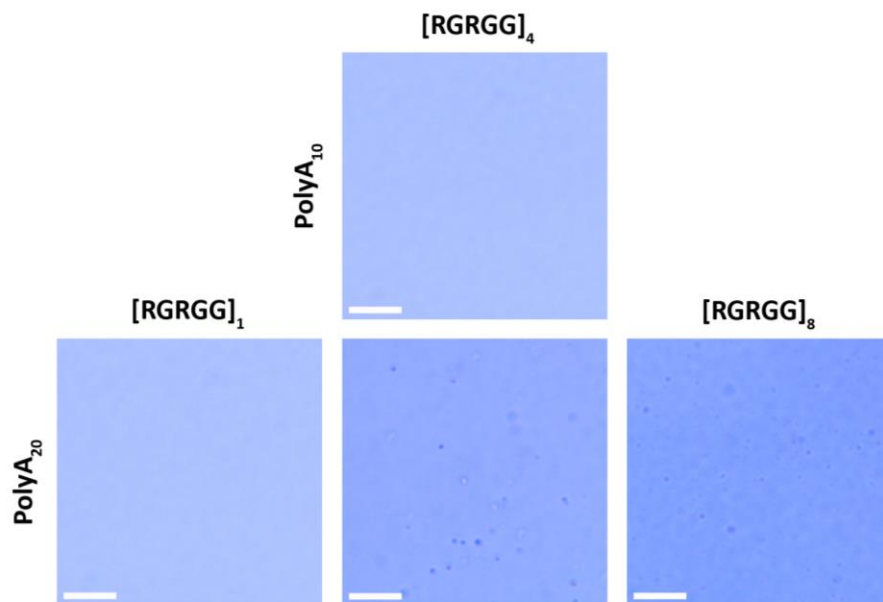


Figure 3.9: Boundary of phase separation for 0.1 mg/ml of peptide and RNA mixtures. Scale bar represents 15 μm . Here the contrast of the images has been adjusted for better visibility of the condensates.

To quantitatively understand the phase separation and its related thermodynamics of the peptide-RNA system we analyzed COCOMO simulations and developed a free-energy thermodynamic model based on enthalpy-entropy decomposition (**Section A2**). Radial distribution functions (RDF) were calculated to extract information on enthalpy for all residue pairs (adenine-adenine, adenine-arginine, adenine-glycine, arginine-arginine, arginine-glycine, and glycine-glycine) in the condensed phase of the simulated systems. We noticed that the RDFs (**Figure 3.10**) do not change substantially with polymer length which indicates a minimal contribution to the length dependence of phase separation by enthalpy. That is, the longer polymers exhibit similar binding strengths as the shorter polymers.

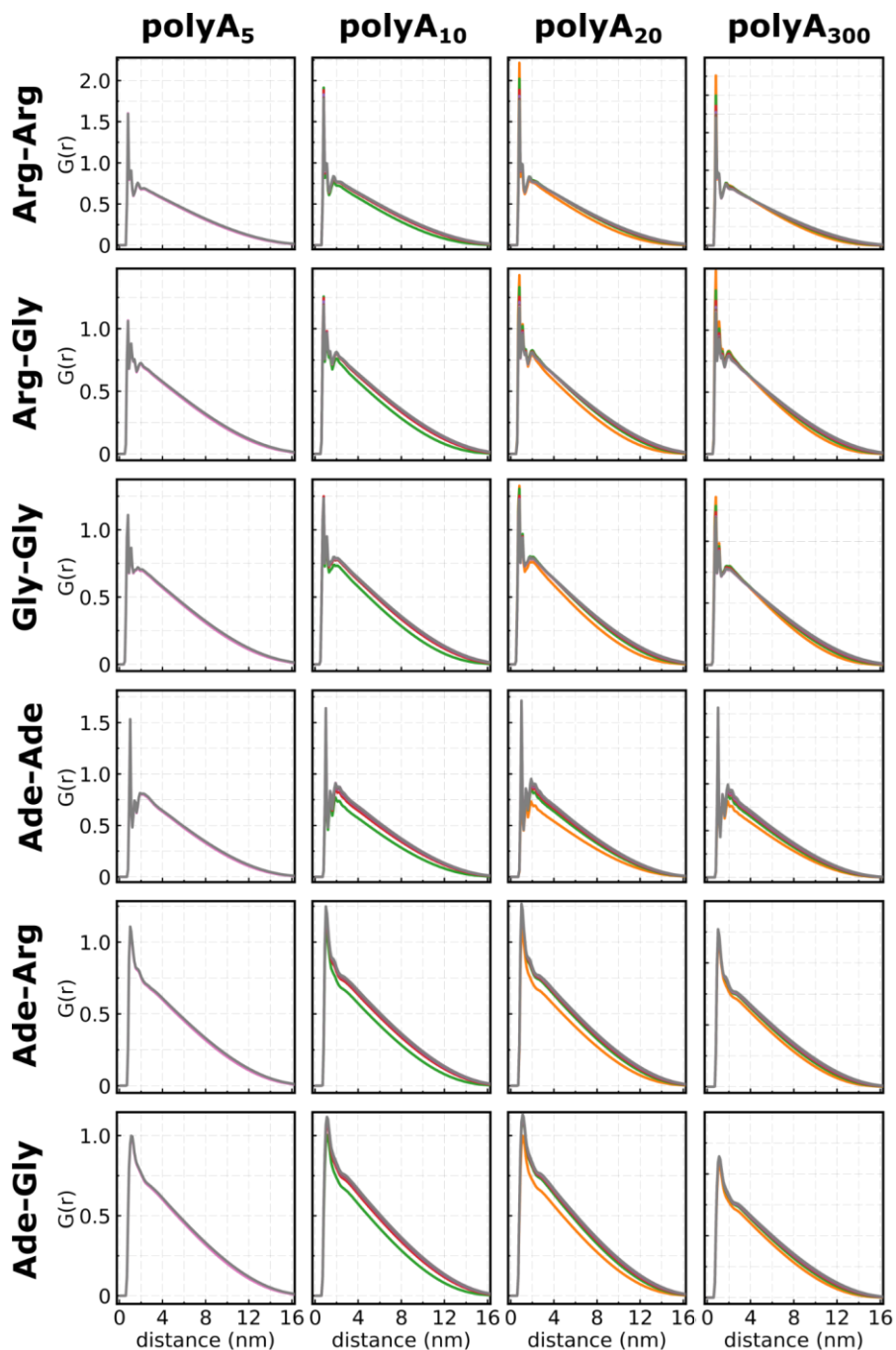


Figure 3.10: Pairwise RDF between different residue types in the condensed phase. Results are shown column-wise for PolyA_N (N = 5, 10, 20 and 300) and row-wise for residue pairs (Arg-Arg, Arg-Gly, Gly-Gly, Ade-Ade, Ade-Arg, and Ade-Gly). Each trace shows a different [RGRGG]_M peptide, where M = 1, 2, 3, 4, 5, 6, 8, and 10 are in blue, orange, green, red, purple, brown, pink and grey respectively. Initial concentrations for these simulations were kept at 1 mg/mL for both protein and RNA.

Enthalpy was calculated using pairwise RDFs, the COCOMO potential and the number of residues, as explained in **Section A2**. Bonded interactions were not considered and intra-chain non-bonded interactions between nearest bonded neighbors were also omitted. Other than that, all inter- and

intra-chain interactions were considered. To obtain the enthalpy of phase separation, the condensed phase was subtracted by the dispersed phase enthalpy. In all systems the enthalpy of the dispersed phase was $\sim 10\%$ of their respective condensed phase and the enthalpy of the coexisting dilute phase was negligibly small, therefore disregarded.

Based on these experimental and simulation results we conclude that the attractive interactions between RNA and peptide are the main drivers of phase separation. But this doesn't explicitly explain how phase separation depends on polymer length. For that we look into the entropy of the system using COCOMO simulations.

3.7 Confinement entropy

Several different entropy contributions were investigated. Loss of translational freedom of the polymers during phase separation or confinement entropy is an obvious contribution. Additionally, conformational changes, counterion effects and entropy arising due to different peptide to RNA molecular ratios between condensed and dispersed phase can contribute as well. To understand these entropic contributions, we analyzed the simulations for information on radii of gyration, probability distributions between residues, peptide/RNA ratios, accessible volumes of the polymers and ion coordination. We realized that the main contributor to entropy is confinement entropy, and the rest of the entropies are not significant (**Section A3**).

Confinement entropy is the change in entropy due to the loss of translational freedom during phase separation. Molecules are restricted to a smaller space inside condensates which will reduce the entropy of the system. We estimated confinement entropy from the ratio of the accessible volume in the condensate to the volume of the box in the simulations using **Eq. A10**. Results are shown in **Figure 3.11**. Here we used the molecular volume of a polymer (Calculated using r_i from **Table A3**) as the accessible volume based on the argument that the high density of the condensates makes the free space too fragmented and creates polymer entanglement restricting most of the translational motion around the position of the polymer itself. Both simulations and experiments showing liquid-like behavior (**Section 3.4**) for the condensates suggest that this is not entirely true, but it is likely a better approximation than assuming that all free volume inside the condensate that

is not occupied by a polymer is accessible. The additional restriction of rotational degrees of freedom was not considered in this analysis.

3.8 Free energy of phase separation

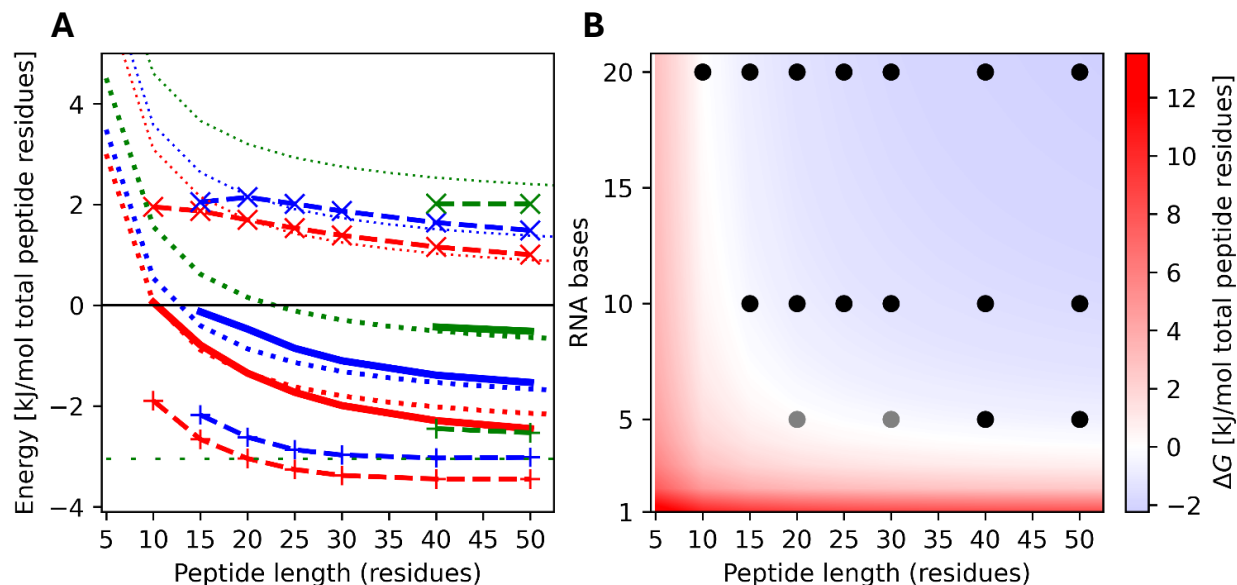


Figure 3.11: Energetic analysis of peptide-RNA phase separation: Here the energy is normalized by the total moles of amino acids in each mixture. **A)** Enthalpies are shown as dashed lines with '+', entropies ($-T\Delta S$ at 300 K) as dashed lines with 'x', and total free energies as solid lines for different RNA (PolyA₂₀: red, PolyA₁₀: blue, PolyA₅: green) as a function of peptide length. Energies were estimated by averaging over five replicate simulations. The statistical errors of the mean are less than 0.1 kJ/mol and are not shown. Short-dash lines reflect total free energy estimates using densities and RDFs based on PolyA₂₀-[RGRGG]₄. Long-dash and dotted lines show the estimated enthalpic and entropic contribution to the free energy, respectively. **B)** The contour plot shows the total free energies as a function of peptide and RNA length obtained with the same parameters. Dots indicate peptide-RNA combinations for which condensates were observed experimentally (black and grey) and in the simulations (black only).

From confocal results we see longer peptide-RNA mixtures phase separating while shorter peptide-RNA mixtures do not (**Figure 3.2**), which is also supported by COCOMO simulations. Since all these mixtures have the same residue (adenine, arginine and glycine) concentrations it is possible that interactions, and therefore enthalpy, are not that different from one mixture to another. This suggests that the observed length dependence of phase separation is dominated by entropy. This claim is supported by the observation of similar droplet sizes for different peptide-RNA mixtures (**Figure 3.5**) indicating the possibility of similar compositions in the condensates and ITC measurements demonstrating similar enthalpies of phase separation regardless of polymer length (**Figure 3.8**). These results give us a qualitative understanding of the investigated system.

Figure 3.11.A shows the calculated enthalpy (Eq. A9), entropy (Eq. A11) and total free energy (Eq. A14) change between the dispersed and condensed phase for each mixture of PolyA_{5,10,20} and [RGRGG]_{1,2,3,4,6,8,10} calculated using COCOMO and the free-energy thermodynamics model (Section A2). Converting the enthalpies to ‘kJ/mol of peptide’ results in values around -100 to -200 kJ/mol of peptide which is in quantitative agreement with ITC results (Figure 3.8). We see that the enthalpy of phase separation (ΔH) is independent of length except for very short polymer mixtures while the change in entropy during phase separation, which is entirely due to the confinement entropy ($\Delta S = -T\Delta S_c$), decreases with length. This is a result of ΔH of the system scaling with the number of residues while ΔS scales with the number of polymers. With increasing polymer length, the condensate size and the number of residues (peptide and RNA) in condensates remain mostly constant in agreement with experiment (Figure 3.5), lowering the number of polymers in the condensate. Therefore, while the entropy is affected by the length dependence the enthalpy remains unchanged.

The free energy of phase separation (ΔG) is negative, indicating the stability of the condensates. Longer polymers form more stable condensates, and stability continuously decreases for shorter lengths. The simulated condensate size and the number of residues in the condensed phase for very short polymers were observed to be slightly less than for longer polymers in contrast to experiment. This difference is reflected in the lower values calculated for ΔH and ΔS . Therefore, to match the experiment by assuming constant condensate sizes and unchanging compositions we set the densities and RDFs of all mixtures to that of PolyA₂₀-[RGRGG]₄ and calculated the energies as indicated in the dotted plots of **Figure 3.11.A**. This increased ΔS , but ΔG was retained which showed that the change in enthalpy at shorter lengths is compensated for by entropy to maintain thermodynamic equilibrium and confinement entropy remains dominant.

Chapter 4

RLP condensate system

This chapter discusses the steps taken to find a suitable protein system and its preparation to investigate the intra-molecular dynamics of IDPs in condensates using the Trp-Cys quenching technique (**Section 2.1**). Several condensate systems were tested to incorporate IDPs into condensates. First, the IDP α -syn^{39w69c} was tested with PEG as a crowder. Here the α -syn is mutated at position 39 and 69 with a Trp and a Cys respectively for Trp-Cys measurements. Condensates were observed at and above 15 % PEG. Trp-Cys measurements indicated a slowdown in protein dynamics in the presence of condensates. Trp-Cys being a bulk measurement detects the condensed as well as the dilute phase. Due to the high viscosity of PEG, the dynamics of α -syn in the dilute phase are slowed down as well. Therefore, it was difficult to distinguish the condensed phase dynamics of α -syn from the dilute phase. Due to these complications, we moved on to a system where the viscosity of the system is independent of the protein concentration. We tested a condensate system of short, oppositely charged peptides, [DFDGD]₆ with [RGRGG]₆ motivated by the length dependent experiments in **Chapter 3**. Here [DFDGD]₆ is mutated with a Trp and a Cys at positions 2 and 29 respectively. Condensates were observed but easily aggregated. Trp-Cys measurements indicated multiple populations which was not easily understood. Therefore, we disregarded this system as well.

Next, we tested the self-condensing resin-like polypeptide SKGP-(GRGDSPYS)₂₀-GY with α -syn. The positively charged RLP acted as a scaffold, forming condensates on its own and incorporating the negatively charged α -syn as a client forming a scaffold-client condensate system. RLP did not add any complexity. Therefore, using this condensate system, we carried out Trp-Cys measurements to calculate the intra-molecular diffusion coefficient of α -syn in the condensed phase, as described in the next chapter.

4.1 RLP sample preparation

RLP was recombinantly expressed as described in **Sections 2.6**. To prepare RLP samples we dissolved 9 mg of lyophilized RLP in 1.5 mL of 50 mM Tris 500 mM NaCl pH 7.4 buffer[29] to get a sample of ~ 350 μ M. The solution was a non-homogeneous mixture with large clusters of

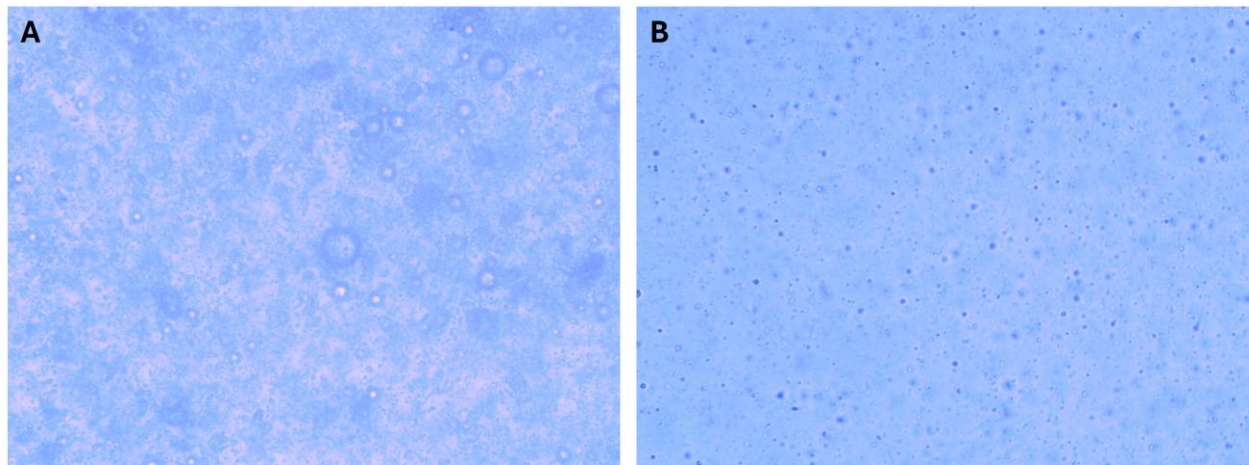


Figure 4.1: Removal of condensate clusters from RLP samples. **A)** Shows large condensate clusters in a freshly dissolved RLP sample. Here 9 mg of RLP was dissolved in 1.5 ml of 50 mM Tris 500 mM NaCl pH 7.4 buffer. **B)** After centrifuging the sample at 4000 rpm for 1 minute and recovering the supernatant the condensates are homogeneous, and the sample is at 170 μ M.

condensates (**Figure 4.1 A**). Therefore, the sample was spun using an Eppendorf centrifuge at 4000 rpm for 1 minute, and the supernatant was recovered which resulted in a homogeneous mixture with individual identifiable condensates as shown in **Figure 4.1 B**.

To perform Trp-Cys quenching measurements a stock solution of 300 μ M of RLP was prepared in 50 mM Tris 500 mM NaCl pH 7.4. 1 mg of RLP was dissolved per 70 μ L of 50 mM Tris 500 mM NaCl pH 7.4 by stirring for about 30 minutes at room temperature and then centrifuged the suspension at 1000 rpm for 1 minute to remove large clusters of condensates and the supernatant was separated. The concentration of RLP was measured using UV spectrophotometry (NanoDrop 2000, Thermo Scientific). Since the sample consists of condensates resulting in UV scattering, measurements were carried out in 6 M Gdn-HCl.

Confocal images were captured using a Nikon A1Rsi Confocal Laser Scanning Microscope. Alexa-488 labeled α -syn39w69c and the Alexa-647 labeled RLP were excited using a diode laser at 488 nm and 647 nm and fluorescence was detected through 525/50 nm and 700/75 nm band-pass emission filters respectively. **Figure 4.2** shows the incorporation of α -syn in RLP condensates captured by confocal microscopy. We see that the self-condensing protein forms condensates at a

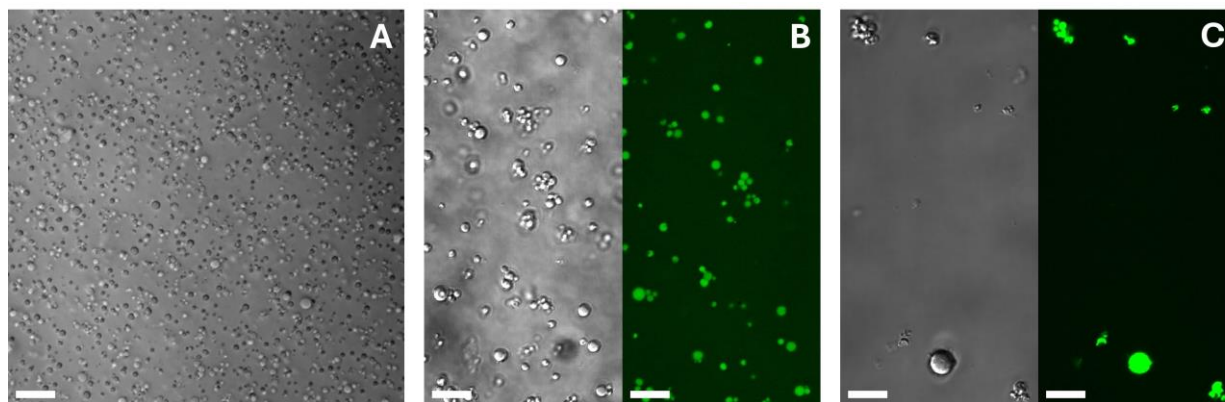


Figure 4.2: Incorporation of α -syn in RLP condensates. Confocal microscopy images captured at 22°C for **A)** 300 μ M RLP in 50 mM Tris 500 mM NaCl pH 7.4 using DIC imaging. DIC on the left and fluorescence on the right for **B)** 10 $\times$ dilution (30 μ M RLP, 50 mM NaCl) and **C)** $\sim$ 40 $\times$ dilution ($\sim$ 8 μ M RLP, $\sim$ 13.33 mM NaCl) of sample **A**. 5 μ M Alexa-488 labeled α -syn is used as the fluorophore, and the samples contain 30 μ M of unlabeled α -syn. Scale bars represent 10 μ m.

size range of $\sim$ 5 μ m down to sub-micron scale in diameter. Addition of Alexa-488 labeled α -syn as the fluorophore shows clear incorporation of α -syn into the condensates.

4.2 Optimizing conditions for Trp-Cys measurements

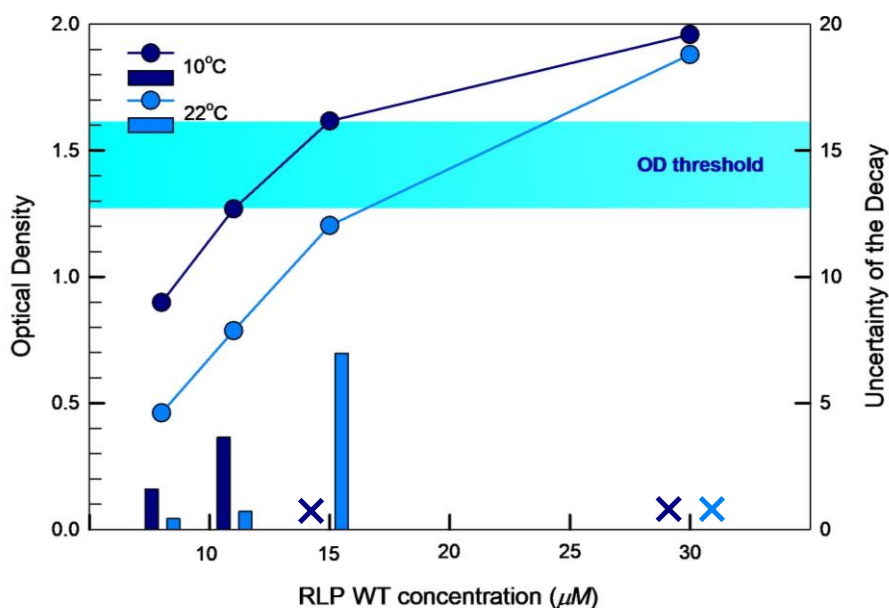


Figure 4.3: Dependence of RLP OD on temperature and RLP concentration. Scatter and line plot represent the ODs of the samples, 50 μ M of α -syn in 8, 11, 15 and 30 μ M of RLP during Trp-Cys measurement at 10 and 22°C. NaCl concentrations are maintained between 10 – 20 mM. OD depends on the optical path length of 1 cm. Above the OD threshold region (cyan band) it is impossible to take measurements due to high turbidity hence indicated by the crosses. Bars represent least square errors of the decay fits.

RLP condensates makes the solution turbid. This was easily observed by the cloudiness of the

samples. When the sample is too turbid there is a significant loss of signal and increased noise due to scattering of the probe laser making it difficult to obtain reliable results from Trp-Cys measurements. The turbidity is sensitive to RLP concentration and temperature. Using the 445 nm probe laser the transmitted light intensity was measured for several concentrations of RLP 8, 11, 15 and 30 μM in a 1 cm cuvette. OD was calculated using $OD = \log(I_0/I)$ with I_0 as the incident and I as the transmitted light. **Figure 4.3** shows the OD and the uncertainty of Trp decay fits for $\alpha\text{-syn39w69c}$ against RLP concentration at 10°C and 22°C. Above OD = 1.6 solutions were too turbid to do a measurement while below OD = 1.3 the solutions were transparent enough. The uncertainties represent the least square errors of the decay fits at the above mentioned RLP concentrations.

Confocal measurements were carried out to identify the saturation concentration of phase separation (C_{sat}) which is defined here as the concentration at which RLP starts phase separating. To identify C_{sat} , RLP concentration was varied from 0.1 μM to 41 μM in 50 mM Tris-HCl 18.33 mM NaCl at pH 7.4 and confocal images (**Figure 4.4**) were captured at room temperature (22°C). 40 nM of Alexa-647 labeled RLP and 30 nM Alexa-488 labeled $\alpha\text{-syn39w69c}$ were used

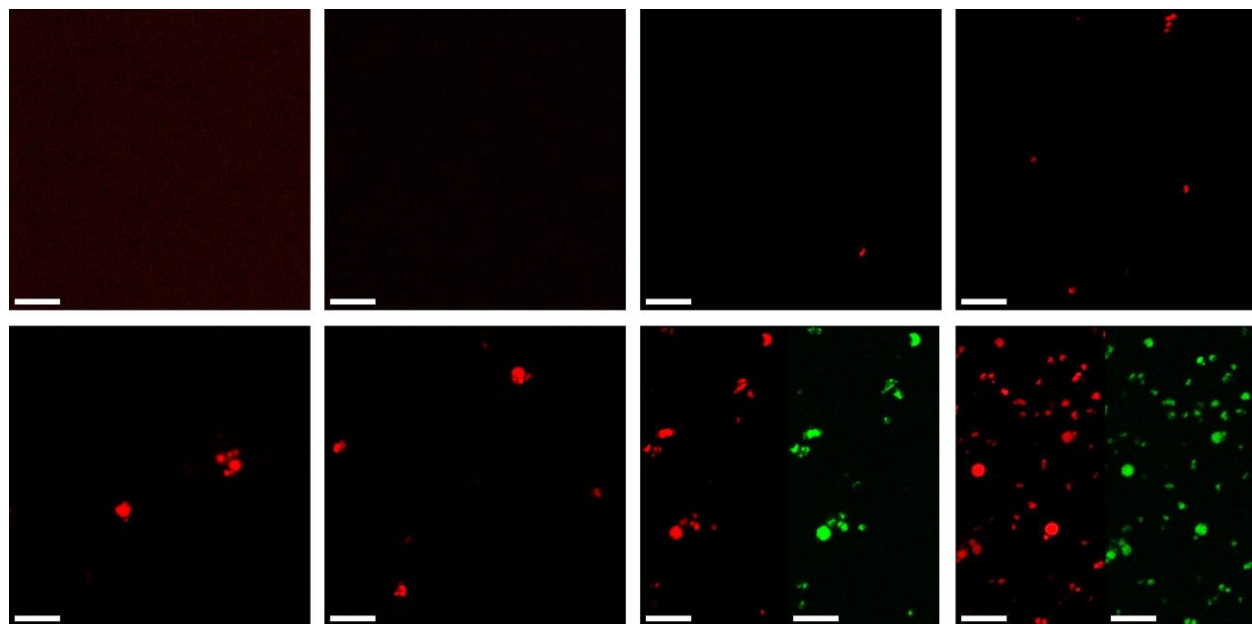


Figure 4.4: C_{sat} of RLP from confocal microscopy. Top row, left to right shows samples of RLP at concentrations of 0.1, 0.5, 0.86 and 3.6 μM and Bottom row, left to right shows RLP at 5.6, 8, 18 and 41 μM in 50 mM Tris-HCl 18.33 mM NaCl pH 7.4. Confocal images were captured at 22°C. Fluorescence is observed with 40 nM Alexa-647 labeled RLP (red) and 30 nM Alexa-488 labeled $\alpha\text{-syn}$ (green). Scale bars are 10 μm .

for fluorescence. Condensates start to disappear below $0.86 \mu\text{M}$. Therefore, we set C_{sat} at $0.86 \mu\text{M}$. For Trp-Cys measurements the concentration of RLP was kept at a low concentration of $8 \mu\text{M}$ above C_{sat} and below the OD threshold region to minimize errors. RLP having a low C_{sat} was an advantage to work at low turbidity.

4.3 Partitioning of α -syn in the condensate solution

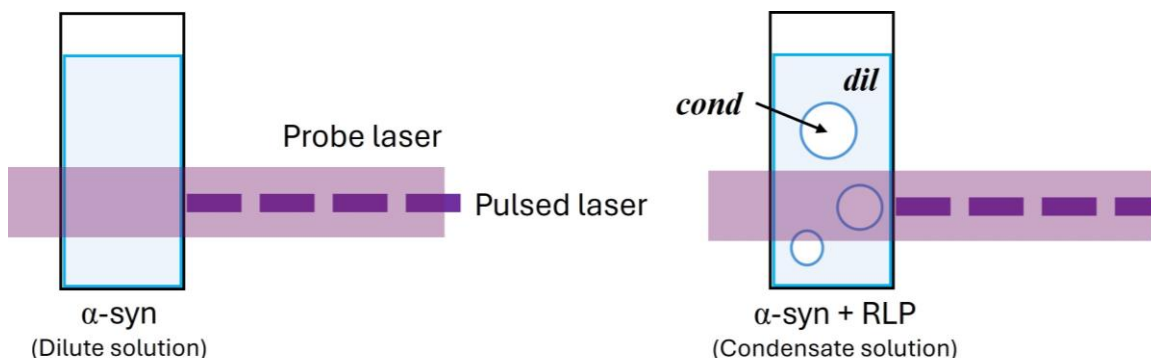


Figure 4.5: Dilute vs condensate samples during Trp-Cys measurements. While the pulsed and probe lasers encounter a single phase for a pure dilute solution (left), with the inclusion of condensates (right) both the dilute and the condensed phases are detected.

Figure 4.5 depicts the encounter of the lasers with a dilute and condensate solution during Trp-Cys measurements. Because the Trp-Cys quenching technique measures the entire sample in bulk, protein in both the dilute and condensed phase of the condensate solution contributes to the observed decay rate proportionate to their relative populations. Therefore, to find the fraction of α -syn in the dilute and the condensed phase we used confocal microscopy. We measured the dilute phase intensity of a sample containing fluorescently labeled α -syn and a sample containing the same concentration of fluorescently labeled α -syn with RLP condensates (**Figure 4.6 A**) and computed their difference. We observe a decrease in intensity with the addition of RLP. This decrease in intensity is proportional to the number of fluorescently labeled α -syn molecules in the condensed phase.

A sample of $20 \mu\text{L}$ of 300 nM Alexa-488 labeled α -syn39w69c was placed on a gridded cover slip and images were captured starting from the coverslip along the z-direction up to a distance of $\sim 3 \mu\text{m}$ at every 125 nm . Similarly, images were captured for the condensate sample of 300 nM Alexa-488 labeled α -syn39w69c in $8 \mu\text{M}$ RLP. **Figure 4.7** shows the images captured. **Figure 4.6 B** shows the dilute phase intensities measured along the z-direction for the dilute and the condensate

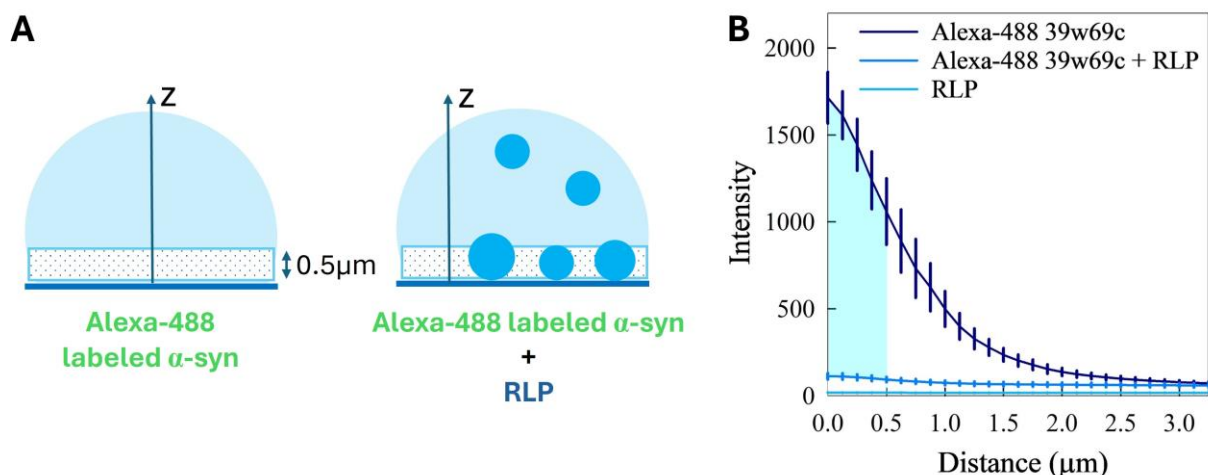


Figure 4.6: Calculating the partitioning of α -syn in a condensate sample using confocal microscopy. A) Measurement of Alexa-488 labeled α -syn sample on a coverslip with (right) and without RLP (left). Images were captured along the z-direction, and the dilute phase intensity was averaged over a distance of 0.5 μ m from the coverslip for both conditions. The intensity profiles obtained along the z-direction are plotted in **B** where the light blue line with error bars represent the condensate sample, dark blue represent the dilute sample and cyan represent RLP alone and serves as the background. All data were averaged over 5 measurements, and the error bars represent the standard error of the mean. The area highlighted in cyan indicates the intensity difference which is proportional to the amount of α -syn that goes into the condensed phase. All samples are in 50 mM Tris-HCl 13.33 mM NaCl pH 7.4 at 22°C (room temperature).

samples by the dark and the light blue plots respectively. The cyan plot shows the background measurement obtained using a sample of 8 μ M RLP with no fluorescent dye present. The background was subtracted from the dilute and condensate sample intensities. The most accurate intensity is measured at the coverslip because the intensity of the highly dispersed fluorophores measured away from the coverslip are attenuated due to optical absorption or scattering from the α -synuclein absorbed on the coverslip. Therefore, the dilute phase intensity difference between the two samples was computed near the coverslip within a distance of 0.5 μ m above it, indicated by the area highlighted in cyan in **Figure 4.6 B**. The computed differences within 0.5 μ m were averaged and divided by the averaged dilute sample intensity within the same region to obtain the fraction of α -syn in the condensed phase, 94 ± 8 %. We assume the same partitioning for the rest of the α -syn mutants.

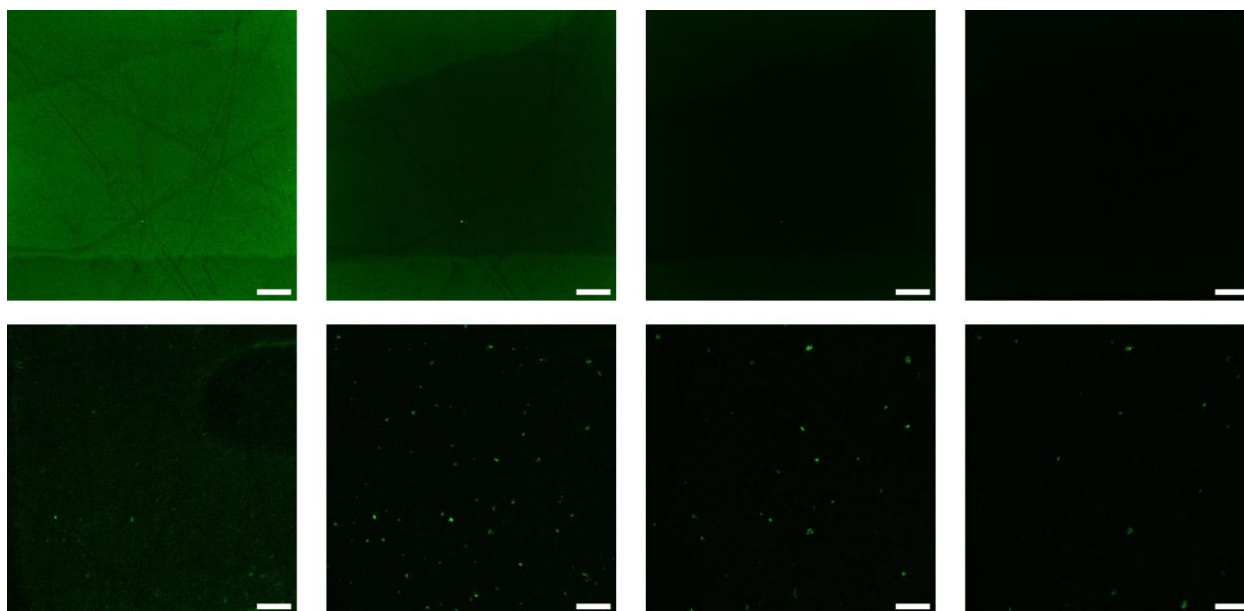


Figure 4.7: Confocal images captured along the z-axis for Alexa-488 labeled α -syn39w69c (top panel) and Alexa-488 labeled α -syn39w69c + RLP (bottom panel) starting from the cover slip surface and 1, 2 and 3 μ m from the surface (left to right). Scale bars represent 15 μ m. Confocal settings were kept constant at laser power = 2 %, offset = 0 and detector gain = 55.

Partition coefficients were calculated for α -syn39w69c and RLP for a sample of 18 μ M RLP 18.33 mM NaCl pH7.4 as shown in **Figure 4.8**. Fluorescent intensity profiles were obtained across condensates and the intensities inside and outside the condensates were averaged separately to obtain the partition coefficients of 84.00 ± 25.22 and 205.74 ± 104.72 for α -syn39w69c and RLP respectively. Measurements were averaged over three droplets. Therefore, a total concentration increase of $\sim 200 - 400$ is observed in condensates relative to the dilute phase.

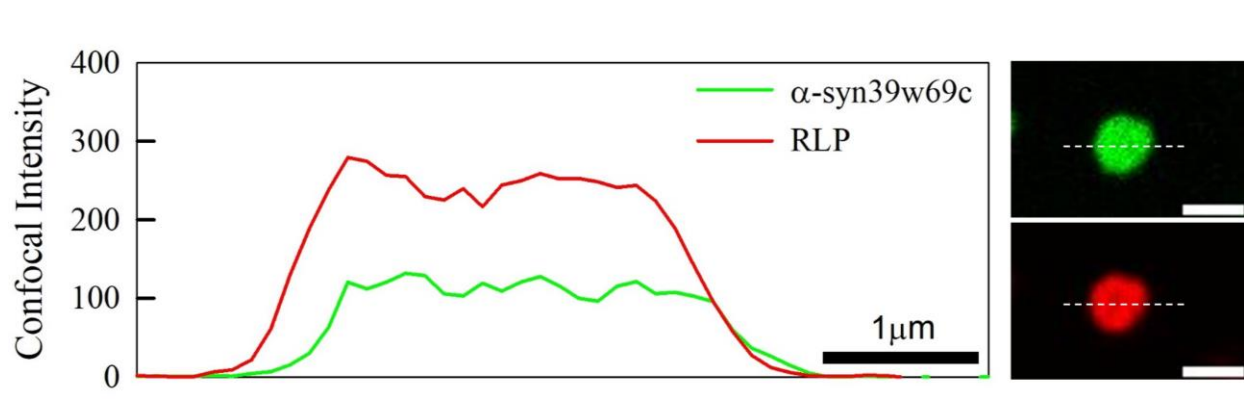


Figure 4.8: Partition coefficients. Plot shows the intensity profiles measured for 30 nM Alexa-488 labeled α -syn39w69c (green) and 40 nM Alexa-647 labeled RLP (red) across an RLP condensate shown on the right. White lines indicate the measurement path. Scale bars are 3 μ m. The intensity profiles have been background subtracted by samples without any fluorophores and normalized by the dilute phase intensity.

Chapter 5

Intra-molecular diffusion of α -syn in condensates

Here we develop the Trp-Cys quenching technique to calculate the intra-molecular diffusion coefficient of α -synuclein, when in a condensate with the aid of COCOMO2 CG molecular dynamic simulations and confocal microscopy.

5.1 α -synuclein dynamics under dilute conditions

α -syn was mutated with a Trp and a Cys placed 25 - 30 residues apart in sequence to carry out Trp-Cys measurements. These mutation sites were selected such that the aggregation propensity of the mutated α -syn is similar to that of the wild-type protein[39]. α -syn39w69c and α -syn69c94w were used to examine chain dynamics unique to different regions of the chain while α -syn-a53t is a Parkinson's Disease causing mutation which slows chain dynamics and α -syn-v74e has shown to increase the dynamic behavior of α -syn39w69c and eliminate aggregation in cells[40, 53]. All α -syn mutants were recombinantly expressed as described in **Section 2.6**.

Trp-Cys measurements were carried out using pump-probe spectroscopy as indicated in **Section 2.3** for 50 μ M α -syn in 50 mM Tris-HCl 13.33 mM NaCl at pH 7.4 and 22°C. α -syn was maintained at a low concentration to prevent inter-molecular quenching between the Trp and the Cys. The observed decays were fit to 1st order exponentials using TableCurve 2D v5.01 to obtain decay rates (k_{obs}). **Figures 5.1 A and D** show the decay time against solution viscosity (μ) for the proteins. The diffusion-limited (k_{D+}) and reaction-limited (k_R) rates extracted from the decay time vs. viscosity plots are depicted in **Figures 5.1 B, E and C, F** respectively. k_{D+} is calculated at the viscosity of water, $\eta = 0.68$ cP.

To determine the Trp-Cys distance distribution ($P(r)$) we carried out molecular dynamic simulations using the COCOMO2 CG model (**Section 2.5**). The initial conformation for a protein chain was randomly assigned such that the chain doesn't overlap with itself using a custom python script. Then each protein chain was placed randomly in a 100 nm size box such that no residue is within 1 nm of each other until the required concentration is reached. All four protein systems were

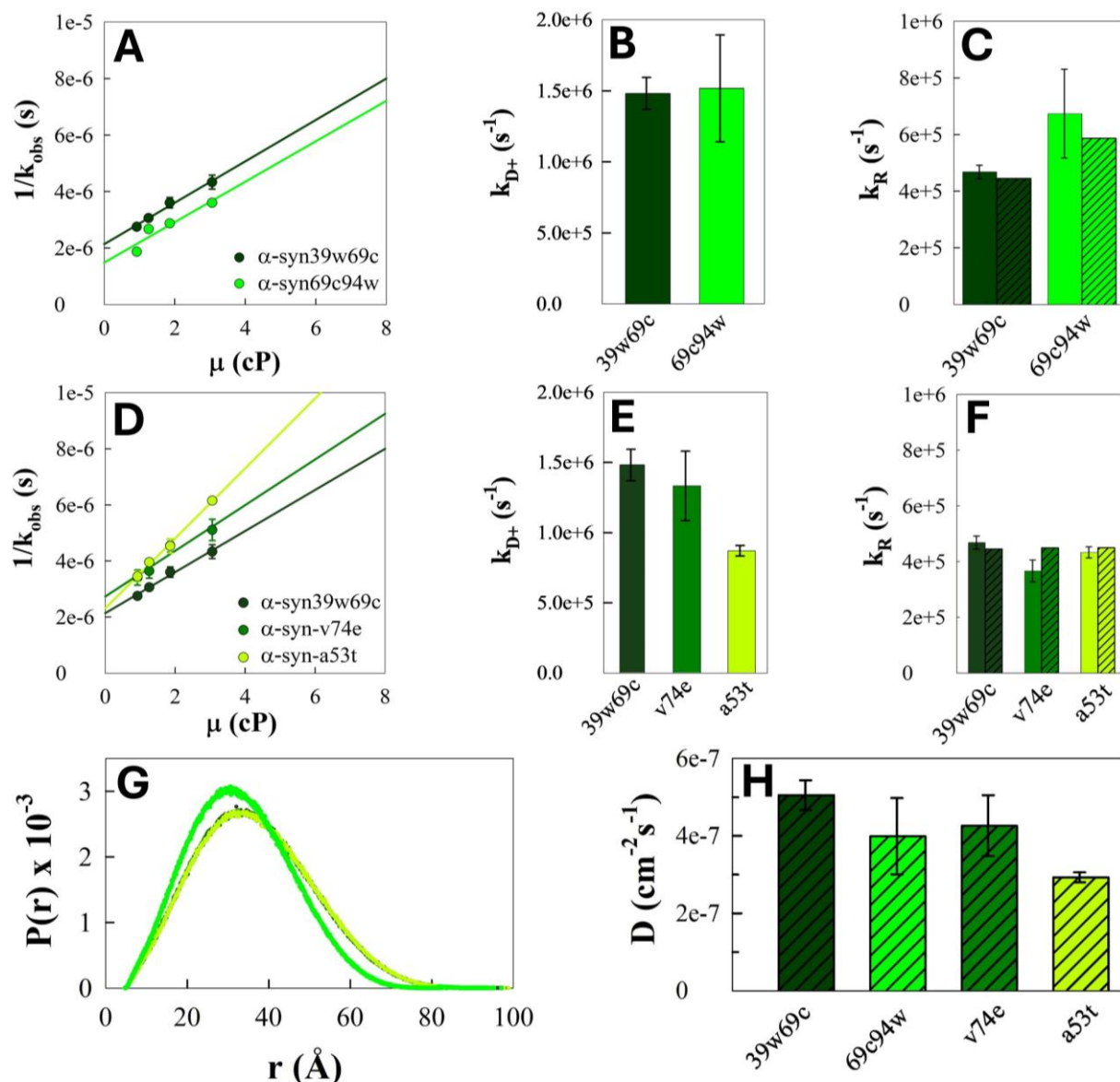


Figure 5.1: Diffusion analysis of α -syn in dilute solution. α -syn39w69c data is compared to α -syn69c94w in **A**, **B** and **C** and to α -syn-v74e and α -syn-a53t in **D**, **E** and **F**. **A** and **D**) Decay time experimentally measured using Trp-Cys quenching technique at viscosities of 0, 10, 20 and 30 % w/w sucrose, 22°C, pH 7.4. The solid bars in **B**, **E** and **C**, **F** shows the calculated diffusion and reaction limited rates respectively from the slopes and intercepts of the decay plots in **A** and **D**. Diagonal striped bars in **C** and **F** represent reaction limited rates calculated using the Trp-Cys distance distributions in **G** obtained from COCOMO2 CG simulations. **H**) Diffusion coefficients of the α -syn mutants.

subjected to CG simulations at 50 μ M, which is the same as the experimental concentrations. Simulations were run for 4 μ s at time steps of 0.02 ps saving every 20 ps. The first 200 ns of the simulation was considered as system equilibration and disregarded. ~ 90 % of the conformations in the systems were monomeric, and multimers were rejected in analysis. Distances between the Trp and the Cys of the same chain were calculated for all molecules in a system for the complete

run and binned by 0.1 Å into a histogram to obtain $P(r)$ shown in **Figure 5.1 G**. Agreement with experimental k_R of α -syn39w69c was obtained using the CG $P(r)$ and **Eq. 2.8** by setting a to 5.25 Å. The rest of the analysis was carried out by fixing a at this determined value. k_R calculated using CG simulations are compared to the experimental values for all four proteins in **Figures 5.1 C** and **F**. The diffusion coefficients of the proteins were calculated using k_R and k_{D+} measured from experiment and $P(r)$ obtained from simulations using **Eq. 2.9** and shown in **Figure 5.1 H**.

The reaction-limited rate of α -syn69c94w obtained from CG simulations is in good agreement with the experimental value (**Figure 5.1 C**) showing an increased rate relative to α -syn39w69c similar to experiment. The chain is more compact compared to α -syn39w69c as depicted by the probability distributions in **Figure 5.1 G** possibly reflecting the nonamyloid component (NAC) region (61-95) hydrophobicity while α -syn39w69c captures the conformations of the N-terminus region. The CG model seems to capture the conformational changes of the chain with good resolution along the chain. In contrast the model is not sensitive enough to capture subtle differences in point mutations. The distance distributions of α -syn-v74e and α -syn-a53t are identical to that of α -syn39w69c as shown in **Figure 5.1 G** and therefore doesn't capture the experimentally measured differences in k_R (**Figure 5.1 F**). Due to the compactness of α -syn69c94w, it is slightly less diffusive, but within error of α -syn39w69c (**Figure 5.1 H**) while α -syn-v74e and α -syn-a53t is mostly diffusion limited. These calculated diffusion coefficients are in agreement with previous studies[3, 39].

5.2 Trp-Cys distance distribution in condensates

To determine the distance distribution between the Trp and the Cys of α -syn in RLP condensates we carried out COCOMO2 CG simulations. Simulations were run with 20 μ M of RLP and 120 μ M of α -syn39w69c starting from monomer configurations in a 100 nm box with 12 molecules of RLP and 72 molecules of α -syn39w69c. Higher concentrations than experiment was used due to low C_{sat} as shown in **Figure 5.2**. However, RLP : α -syn39w69c concentration ratio was maintained at 1 : 6 similar to experiment. Initial random placement of the chains were performed similarly to dilute phase simulations in **Section 5.1**. Run time was 4 μ s with a time step of 0.02 ps and coordinates were saved every 20 ps. Nucleation of condensates were observed $\sim$ 1.5 μ s and the

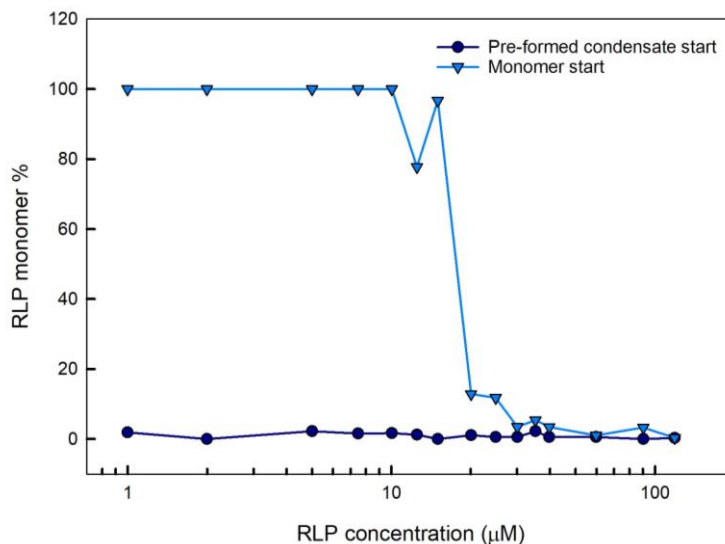


Figure 5.2: C_{sat} of RLP for different starting conditions in CG simulation. Fractions of RLP monomers in the dilute phase is plotted against the total RLP concentration. Starting with a system of randomly placed monomers the system takes longer to equilibrate and overestimates $C_{sat} \sim 20 \mu\text{M}$. Starting with a pre-formed condensate no dissolution of the condensate is observed even at concentrations as low as $1 \mu\text{M}$ indicating that C_{sat} probably lies below $1 \mu\text{M}$. CG simulations were performed at different concentrations of RLP from $1 - 120 \mu\text{M}$ to observe. 180 molecules of RLP were simulated for $5 \mu\text{s}$ and the monomer fractions were averaged over the last μs . The concentrations of the systems were controlled by changing the box size. The data was obtained from [2].

system comes to an equilibrium $\sim 2.5 \mu\text{s}$ (**Figure A7 B**). $P(r)$ in the condensate was calculated during the last $1.4 \mu\text{s}$ of the run. k_R in the condensed phase (**Table 5.1**) was calculated using **Eq.2.8**. Here $a = 5.25 \text{ \AA}$ which was determined by dilute solution Trp-Cys measurements for the CG model (**Section 5.1**). COCOMO2 CG simulations were carried out for the other three α -syn mutants at $120 \mu\text{M}$ with RLP at $20 \mu\text{M}$ to acquire phase separation. The obtained $P(r)$ are shown in **Figure 5.3**. α -syn-v74e did not phase-separate with RLP.

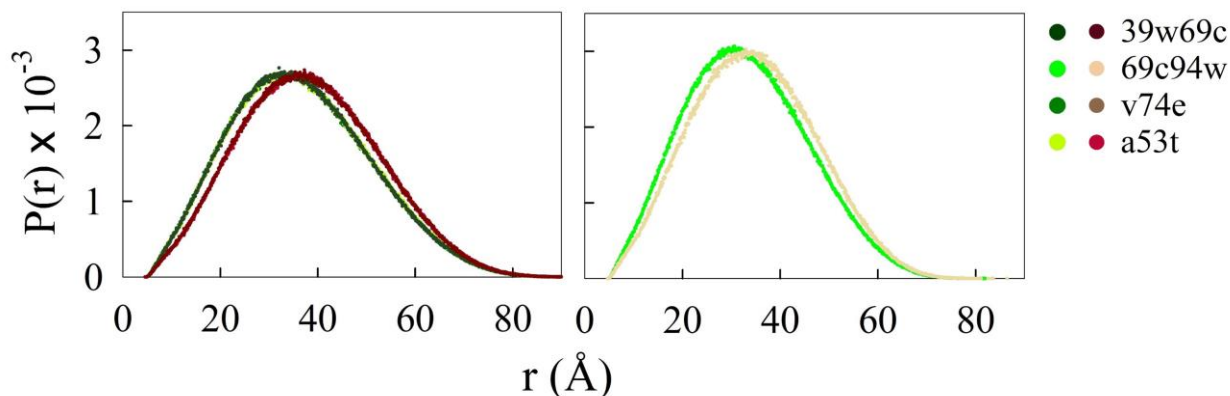


Figure 5.3: $P(r)$ in the condensed phase. Comparison of Trp-Cys distance distributions in the condensed phase (represented in red shades) to dilute phase (represented in green shades) for α -syn39w69c, α -syn-v74e and α -syn-a53t on the left and α -syn69c94w on the right.

To exactly match the experimental conditions simulations were run with a pre-formed RLP condensate with 8 μM of RLP and 50 μM of $\alpha\text{-syn}39\text{w}69\text{c}$ by the Feig group. Starting from a pre-formed condensate shows a $C_{\text{sat}} < 1 \mu\text{M}$ (**Figure 5.2**) similar to experiment. The pre-formed condensate was initially formed at a high concentration, 80 molecules of RLP in a 100 nm box and then the condensate was transferred to a 255 nm box together with 500 molecules of $\alpha\text{-syn}39\text{w}69\text{c}$. The system was simulated for 10 μs with a step size of 1 ns and coordinates were saved every 10 ns. System equilibrates $\sim 5 \mu\text{s}$. $P(r)$ in the condensate was calculated during the last 5 μs of the run which was extremely similar to the monomer start $P(r)$ (**Figure A7 A**). However, due to poor sampling as a result of computational limitations the $P(r)$ computed from the pre-formed condensate start system was less accurate than the monomer start $P(r)$ therefore, not used.

5.3 Inter-molecular quenching in condensates

The protein concentration inside a condensate is $\sim 200 - 400$ times higher than the dilute phase, which could result in the Trp being quenched by other sources instead of the Cys on the same chain. The most likely quencher besides Cys is Tyrosine (Tyr)[54]. There are 4 Tyr in $\alpha\text{-syn}$ and 21 Tyr in RLP. Therefore, we carried out Trp quenching measurements for the dilute and condensate solutions with $\alpha\text{-syn}94\text{w}$ which lacks the Cys (**Figure 5.4 A**). The Trp decay rate was measured for 50 μM $\alpha\text{-syn}94\text{w}$ in 50 mM Tris-HCl pH 7.4 at 22°C in a dilute solution. The decay

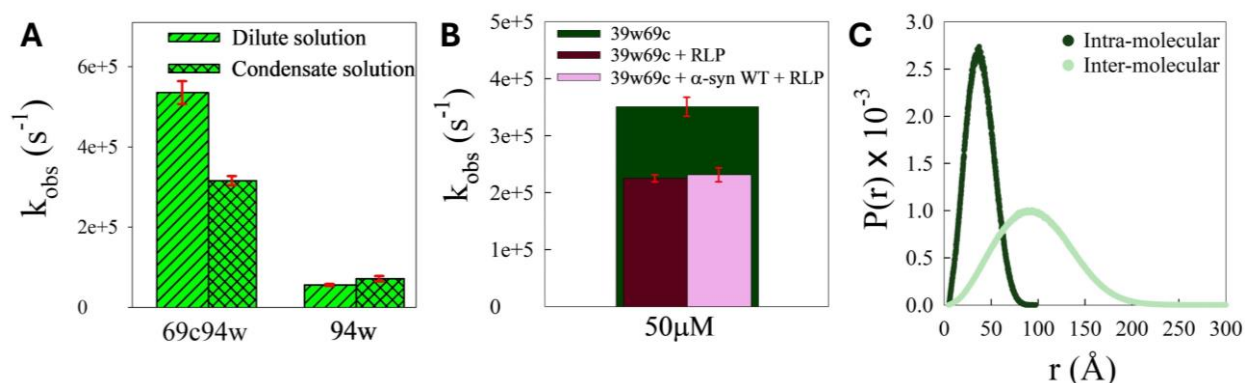


Figure 5.4: Crowding in condensates. **A)** Comparison of $\alpha\text{-syn}$ Trp decay with and without the quencher (Cys) in the dilute (diagonal striped bars) and the condensate solutions (checked bars) to investigate Tyr quenching. Dynamics are very similar for $\alpha\text{-syn}94\text{w}$ in the dilute and condensate solutions indicating the minimal effect of Tyr quenching. **B)** Test to identify inter-molecular quenching by the Cys in crowded conditions. 40 % of the 50 μM $\alpha\text{-syn}39\text{w}69\text{c}$ (decay rate shown in red) molecules in a condensate solution were replaced by 20 μM of $\alpha\text{-syn WT}$ (decay rate shown in lilac). No change within error indicates the absence of inter-molecular quenching. **C)** Intra- (dark green) and inter-molecular (light green) distance distributions of $\alpha\text{-syn}39\text{w}69\text{c}$ in the condensed phase obtained from CG simulations. The number of Trp and Cys encounters at close contact ($< 10 \text{\AA}$) on different chains are significantly less compared to same chain encounters for inter-molecular quenching to be considered.

is in the range of the natural decay of Trp (20 μ s for 70 μ M NATA at room temperature[35]) showing that the intra-molecular quenching effect from the 4 Tyr on the α -syn chain has a minimal effect compared to Cys quenching. With the addition of 8 μ M of RLP we see only a slight increase in the decay rate for α -syn94w and almost within error which indicates that the large number of Tyr near a Trp due to crowding does not affect quenching.

Another source of quenching is a Cys from another chain. To investigate this, we measured the Trp decay of α -syn39w69c with 8 μ M RLP in 50 mM Tris-HCl 13.33 mM NaCl pH 7.4 at 22°C by replacing 20 μ M of the 50 μ M α -syn39w69c with α -syn WT which does not contain either Trp or Cys. The total concentrations of the proteins were kept unchanged to retain the same crowding effect but the concentration of Cys was reduced by 40 %. The decay rate shows no change (**Figure 5.4 B**) which indicates that the Cys of other chains do not diffuse close enough to quench Trp. This observation is also supported by CG simulations as shown in **Figure 5.4 C** where we compare the inter-molecular P(r) between Trp and Cys to intra-molecular P(r). The inter-molecular distribution is shifted to longer lengths. At distances less than 10 Å where the quenching is significant the probability of a Cys from a different chain interacting with the Trp is ~ 15 times less than the interaction with the Cys from its own chain. These results indicate that Trp-Cys intra-molecular quenching dominates the observed decay in the condensates.

5.4 Translational diffusion of α -syn in condensates

Fluorescence recovery after photo bleaching (FRAP) measurements were performed for 300 nM Alexa-488 labeled α -syn39w69c in 30 μ M RLP condensates using an Olympus FV1000 confocal microscope at 23°C. Sample was prepared in 50 mM Tris 50 mM NaCl at pH 7.4. Alexa-488 labeled α -syn39w69c was excited using a diode laser at 488 nm and fluorescence was detected through a 520/50 nm band-pass emission filter. A circular area of radius $r = 0.5 \mu$ m was bleached to 1/3 of initial intensity for 100 ns at the center of a selected condensate and full recovery of the bleached area was reached within 38 s indicating the liquid nature of the condensates as shown in **Figure 5.5**. Fluorescence recovery data was fit with[55],

$$f(t) = a \times \exp(-2\tau_R/t) [I_0(2\tau_D/t) + I_1(2\tau_R/t)] \quad \text{Eq. 5.1}$$

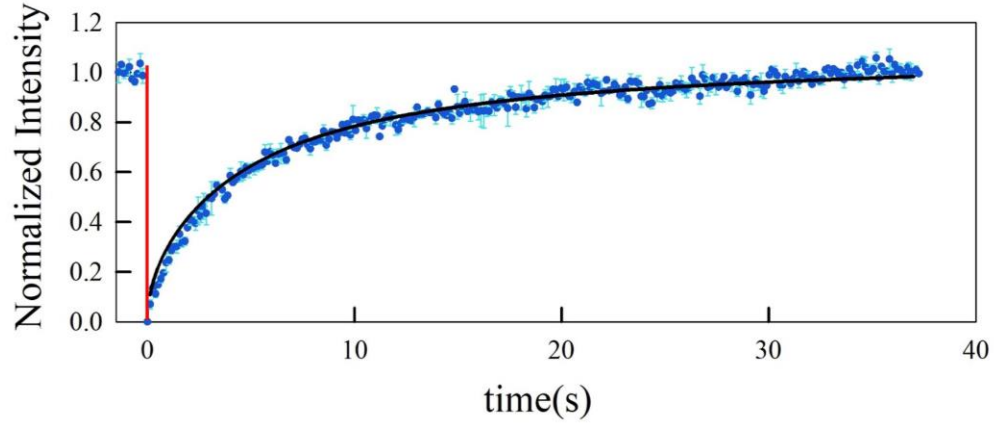


Figure 5.5: FRAP of α -syn39w69c in RLP condensates. Fluorescence intensity of the bleached area vs time is plotted in the upper panel. Intensity is normalized by the average intensity before bleaching. Bleaching event is indicated by the red vertical line. Confocal images are shown in the order of pre-bleaching (< 0 s), bleaching (0 s) and post-bleaching (> 0 s) in the lower panel. Scale bar: 5 μ m.

which assumes 2D radial diffusion resulting in a fluorescence recovery time of $\tau_R = 3.97$ s. Here I_0 and I_1 are modified Bessel functions and a is a proportionality constant. The apparent translational diffusion was calculated by,

$$D_{app} = \frac{r^2}{4\tau_R} \quad \text{Eq. 5.2}$$

where r^2 is the mean squared distance travelled by a fluorophore during recovery resulting in $D_{app} = 0.0016 \text{ Å}^2\text{ns}^{-1}$. This result is in agreement with typical translational diffusion observed for condensates[4, 28, 29].

5.5 Intra-molecular diffusion of α -syn in condensates

The typical calculation of the intra-molecular diffusion coefficients uses the reaction and diffusion limited rates obtained from the linear relationship between the decay time and the viscosity of the sample obtained experimentally as indicated for dilute solutions in **Figure 5.1** where the viscosity is controlled using sucrose. However, we observed that RLP condensates are significantly reduced

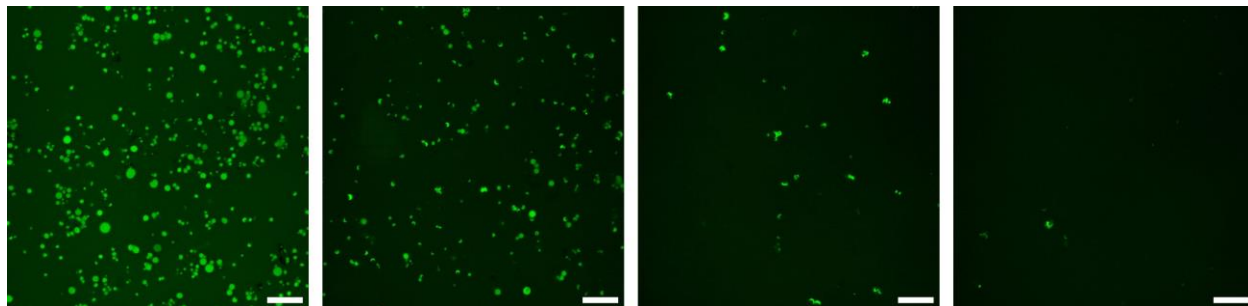


Figure 5.6: High sucrose concentrations dissolve α -syn + RLP condensates. Confocal microscopy images of 30 μ M α -syn39w69c in 32.8 μ M RLP in the presence of 0, 10, 20 and 30 % w/w sucrose (from left to right). Solvent is 50 mM Tris-HCl 50 mM NaCl pH 7.4. 5 μ M of Alexa-488 labeled α -syn39w69c is used for fluorescence. Images were captured at 22°C. Scale bars: 10 μ m.

with the addition of sucrose as shown in **Figure 5.6**. To overcome this limitation, we calculated k_R from CG simulations (**Section 5.2**).

Trp-Cys quenching measurements were performed for 50 μ M α -syn39w69c, α -syn69c94w, α -syn-v74e and α -syn-a53t with 8 μ M RLP using pump-probe spectroscopy at 22°C and pH 7.4 in 50 mM Tris-HCl 13.33 mM NaCl as indicated in **Section 2.3**. Measured decay rates (k_{obs}) are shown in **Figure 5.7**. We observe a significant decrease in rates with the presence of condensates relative to the dilute solutions indicating slower dynamics. Further analysis of the decay transition from dilute to condensed phase is shown in **Section A3**. By considering k_{obs} as a linear combination between the dilute phase (k_{dil}) and the condensed phase (k_{cond}) decay rates as explained in **Section 4.3** we write,

$$k_{obs} = ak_{dil} + bk_{cond} \quad \text{Eq. 5.3}$$

where, $a = 0.06 \pm 0.01$ and $b = 0.94 \pm 0.08$ are the fraction of α -syn in the dilute and the condensed phase respectively as measured by confocal microscopy (**Section 4.3**). We assume that k_{dil} is equal to the decay rate in a pure dilute solution shown in **Figure 5.7**. Therefore from **Eq. 5.3** we calculate the decay rates in the condensed phase k_{cond} , which are shown in **Table 5.1** for all four mutants

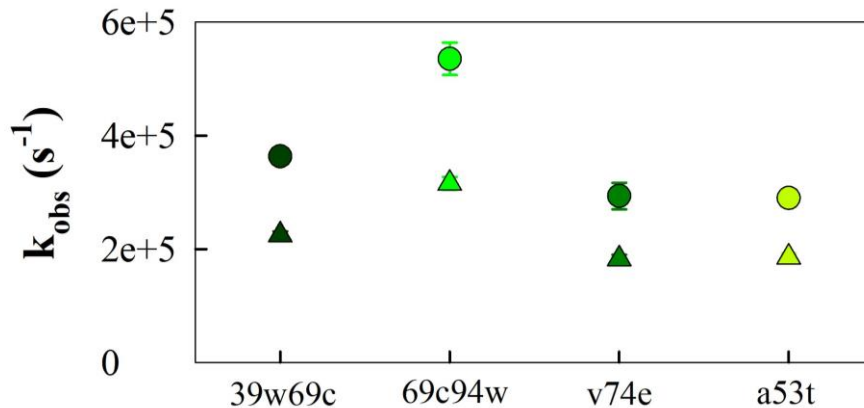


Figure 5.7: Decay rates in dilute vs condensate solutions. Measured Trp decay rates for the α -syn mutants in the pure dilute (circle) and condensate (triangle) solutions at 22°C. Dilute solution data are from **Section 5.1**. Different shades of green represent the α -syn mutants similarly to **Figure 5.1**.

of α -syn. Using **Eq. 2.6**, k_{cond} and k_R obtained from CG simulations, we calculate k_{D+} . The intra-molecular diffusion coefficient was calculated using **Eq. 2.9**. **Table 5.1** shows the computed k_{D+} and D for all α -syn mutants.

Intra-molecular diffusion coefficients of the α -syn mutants in the condensed phase are compared to the dilute solution values in **Figure 5.8**. We do not observe phase separation for α -syn-v74e in CG simulations therefore we are unable to calculate D . We observe that diffusion is reduced by $\sim 1/2$ for all the proteins when in condensates. Slowdown of protein motion in the condensed phase is to be expected because the highly concentrated crowded environment can result in higher

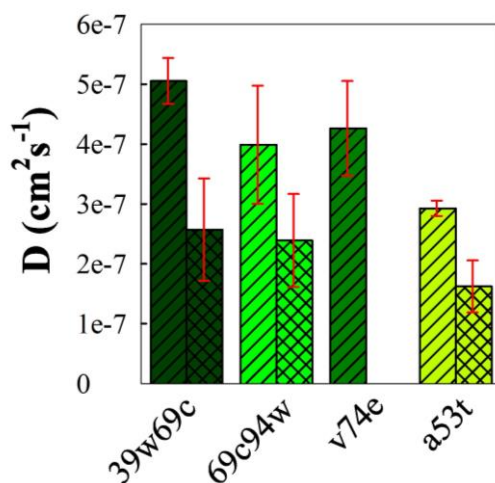


Figure 5.8: Intra-molecular diffusion coefficients of α -syn mutants in the condensed phase. Comparison of the diffusion coefficients in the condensed phase (checked bars) to the dilute solution (diagonal striped bars). Different shades of green represent the α -syn mutants similar to **Figure 5.1**.

viscosities and frequent and stronger transient interactions. However, we calculate condensate concentrations ~ 300 times higher than the dilute phase. These high concentrations and the D_{app} calculated for α -syn in RLP condensates using FRAP are in good agreement with measurements carried out for a similar system which have shown translational diffusion rates ~ 30 times lower relative to the dilute phase and viscosities and concentrations ~ 300 times and ~ 1000 higher in the condensed phase while showing only ~ 3 times decrease of protein reconfiguration rates under crowded conditions[4]. They claim the highly dynamic behavior to be a result of rapid switching of pico- to nanosecond transient contacts. Given the high concentration of our system, we however observe fast dynamics for α -syn in condensates. To understand this, we look into protein interactions from CG simulations.

Table 5.1: Diffusion coefficient in the condensed phase for the α -syn mutants.

Protein	$k_{cond}(s^{-1})$	$k_R(s^{-1})$	$k_{D+}(s^{-1})$	$D(cm^{-2}s^{-1})$ $\times 10^{-7}$
α -syn39w69c	216233.91 ± 26266.04	340630	592106.70 ± 196945.92	2.57 ± 0.86
α -syn69c94w	301890.63 ± 39905.96	508360	743302.14 ± 241918.62	2.39 ± 0.78
α -syn-v74e	175743.92 ± 23953.24			
α -syn-a53t	179307.75 ± 23438.17	349740	367953.24 ± 98698.56	1.62 ± 0.44

Residue-residue contact frequency maps from CG simulations for α -syn39w69c and RLP are shown in **Figure 5.9**. Interactions between α -syn molecules show an overall higher frequency in the condensed phase compared to the dilute solution due to the highly crowded environment but the interacting sites appear to be the same (**Figure 5.9 A and B**). In the condensed phase (**Figure 5.9 C, D and E**) interactions are more prominent between RLP molecules than α -syn. RLP molecules seem to interact with each other preferably through the central region of the chain which consists of net charge neutral segments. The negative charged C-terminal of α -syn interacts electrostatically with the positively charged Lysine at the N-terminal of RLP which can be the dominant interaction that recruits α -syn into RLP condensates. Overall, interactions are observed to be electrostatically driven. Therefore, the observed fast intra-molecular dynamics for α -syn in condensates given their high concentration and possibility of high viscosities and slow translational diffusion can be due to less frequent interactions between α -syn- α -syn and α -syn-RLP

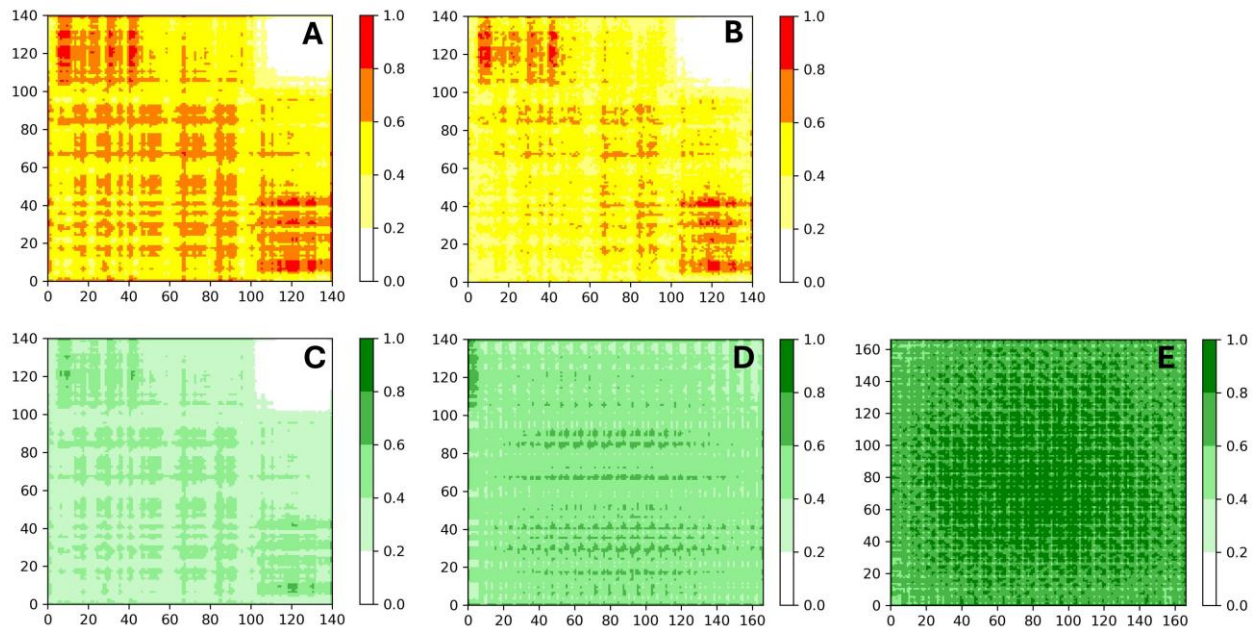


Figure 5.9: Residue-residue contact frequencies between proteins for α -syn39w69c. α -syn- α -syn in the **A)** condensed phase and **B)** dilute solution normalized by their maximum frequency value. **C)** α -syn- α -syn **D)** α -syn-RLP and **E)** RLP-RLP in the condensed phase normalized by their total possible contacts per residue pair. The contact cutoff was set at 7 Å. Color-bars indicate frequency. More frequent contacts are indicated in darker colors.

compared to highly frequent RLP-RLP interactions. In addition, the conformations are only slightly extended in condensates relative to the dilute solution as shown in **Figure 5.3**. To understand this behavior of α -syn in the condensed phase we are pursuing atomistic simulations. Preliminary data has shown good agreement with experiment. A detailed study of atomistic simulations will provide information on the faster dynamics of α -syn and underlying mechanisms.

Chapter 6

Characterization of protein-protein interactions in a crowded system

Here we study transient protein-protein interactions under crowded conditions. We utilize the Trp-Cys quenching technique (Section 2.1) as an experimental approach to characterize these interactions. The technique is sensitive to sub-nanometer lengths, making it suitable for investigating transient interactions between residues at very close contact under crowded conditions. Here we mutate sequence variants of the HP35 fragment of the Villin headpiece and the T22G mutant of the drkN SH3 domain with a Trp which acts as the probe and the B1 domain of Protein G with a Cys (Table 6.1) which acts as the quencher and crowder. The experimental Trp-Cys quenching was supported by molecular dynamics simulations. All simulations were performed by the Feig group.

6.1 Trp-Cys protein-protein quenching

Proteins (Protein G, Villin wild-type (WT), Villin mutants: V10W, K33W, R15T+K30E (double-mutant) and SH3) were purchased from Biosynthesis in the form of lyophilized powders. Initial stock of protein G was prepared by dissolving in 50 mM HEPES buffer at pH 7.0 and sonicating for 1 hr. 300 μ M, 300 μ L aliquots of the peptides were prepared by dissolving in the HEPES buffer. Both protein G and peptide stocks were stored at -20°C . Prior to measurements Protein G was thawed and 2 mL of it was degassed using N_2O (Section 2.3) for 1 hr in a sealed vial and then transferred into a degassed and sealed long neck quartz cuvette (Hellma) using an airtight syringe (Hamilton, GASTIGHT) after which an aliquot of the peptide was thawed and added to the cuvette.

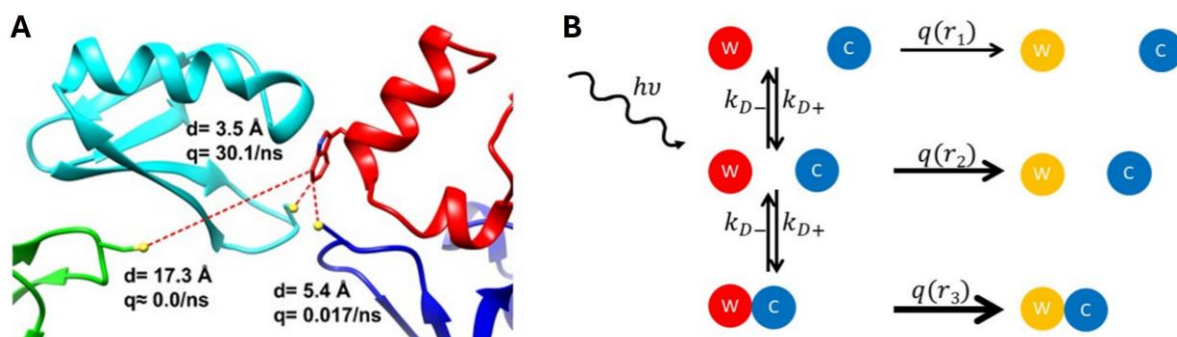


Figure 6.1: Kinetic model for the Villin/SH3-ProteinG crowded system. A) Villin Trp is surrounded by multiple Protein G Cys. A Trp has a higher probability to encounter a Cys due to their abundance hence speeding up decay. B) Kinetic model for the distance dependent quenching of the Trp by multiple Cys under crowded conditions.

Measurements were carried out using pump probe spectroscopy (**Section 2.3**) at 20°C. They were obtained within ~ 15 minutes of addition of Villin/SH3 to Protein G and averaged over 6 repeats per sample.

Trp-Cys measurements were carried out for systems of Trp mutated Villin variants and SH3 at a concentration of 40 μ M with Cys mutated Protein G at a high concentration. Protein G quenches the Trp triplet state upon close contact with Villin/SH3. Unlike in a dilute solution (**Figure 2.1**), under crowded conditions a Trp is surrounded by multiple Cys which results in complicated distance dependent kinetics (**Figure 6.1**). Therefore, the experiments are paired with computational modeling of fully atomistic and CG molecular dynamics simulations for interpretation of the experimental observables and to gain molecular-level insights.

Figure 6.2.A shows the obtained decays and their derivatives in log scale for Villin K33W in the presence of the quencher Protein G at different concentrations ranging from zero to 840 μ M. With

Table 6.1. Amino acid sequences and electrostatic charge of the proteins. The first methionine residue (highlighted in green) was not present in the experimental constructs of Villin. Tryptophan residues and point mutations are highlighted in blue and red, respectively.

Protein	Charge	Sequence
Villin WT (N28H)	+2	MLSDEDFKAV FGMTRSAFAN LPLWKQQHLK KEKGLF
Villin V10W	+2	MLSDEDFKAW FGMTRSAFAN LPLYKQQNLK KEKGLF
Villin K33W	+1	MLSDEDFKAV FGMTRSAFAN LPLYKQQNLK KEWGLF
Villin R15T+K30E (N28H)	-1	MLSDEDFKAV FGMTTSAFAN LPLWKQQHLE KEKGLF
drkN SH3 (T22G)	-4	MEAIKHDFFS ATADDELSFR KGQILKILNM EDDSNWYRAE LDGKEGLIPS NYIEMKNHD
Protein G (K10C W43Y)	-5	MTYKLILNGC TLKGETTTEA VDAATAEKVF KQYANDNGVD GEYTYDDATK TFTVTE

no crowder the Trp exhibits its unquenched lifetime of $\sim 40 \mu\text{s}$ [35]. With the addition of Protein G we see faster decays due to Cys quenching, more clearly distinguishable in the derivative plots. With the increase in Protein G concentration the location of the derivative minima and its depth changes. The minima shift to shorter times, indicating more frequent protein-protein contacts due to a higher concentration of Cys near the Trp while the increase in the minima depth probably indicates the shift from slow intra-molecular quenching by other amino acids in the Villin and nearby water molecules to more efficient inter-molecular quenching by the Cys. The nonexponential decay observed in these results is due to low probability quenching by non-Cys amino acids within the folded structure[35]. **Figure 6.2.B** and **C** shows Trp-Cys measurements of Villin variants and SH3 respectively in the presence of the Protein G crowder at $840 \mu\text{M}$. The Villin variants do not seem to shift in time but do exhibit different derivative depths which suggests that we can probe differences in local interactions with respect to where Trp is located for each variant (**Table 6.1**). The decay observed for SH3 is much slower than Villin indicating that the

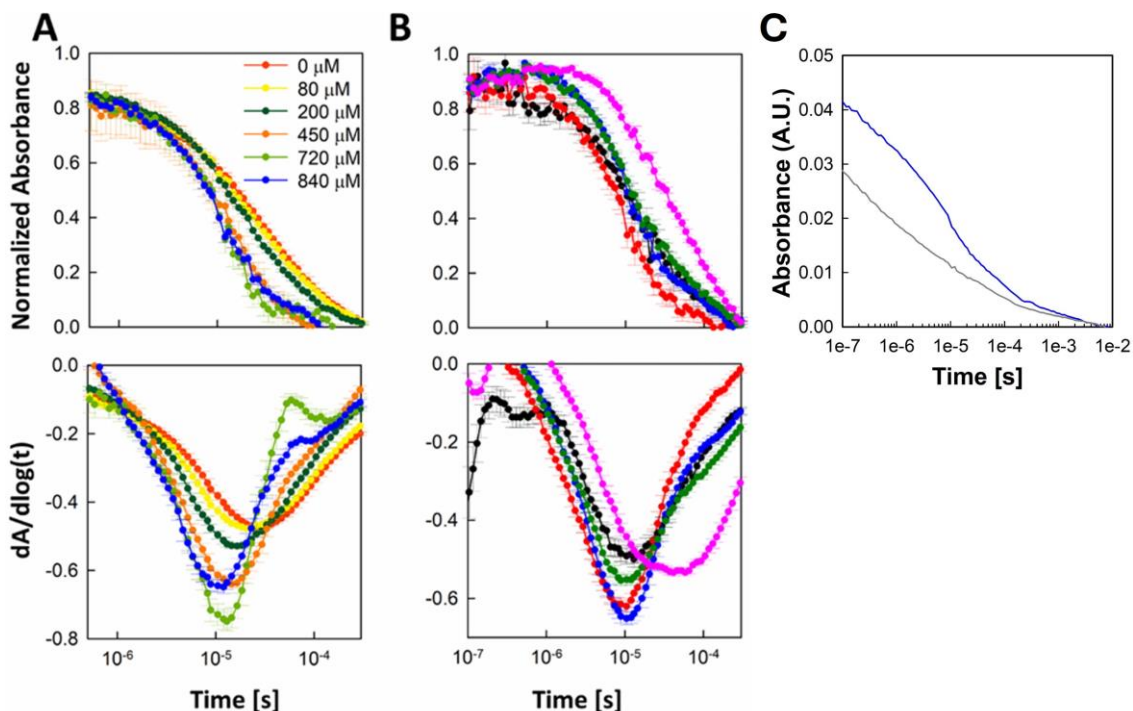


Figure 6.2: Experimental measurements of Trp triplet lifetime in Villin variants and SH3 in the presence of Protein G. Decays are shown as absorbance vs time (top) and as the derivatives of the decays vs $\log(t)$ calculated from the slope of a linear fit over a range of 21 time points (bottom). **A)** Villin K33W at various concentrations of protein G. **B)** 40 μM Villin WT (black), V10W (red), K33W (blue), R15T+K30E (green), and SH3 (magenta) in the presence of 840 μM Protein G. Curves were averaged from six independent measurements and normalized to 0.88 at 147 ns and to 0 at 369 μs . Error bars represent the standard deviation. Error bars in the derivatives represent the error of the fit. **C)** The Protein G background (grey) is subtracted from Protein G + Villin K33W (blue) Trp-Cys decay measurement to obtain the data in **B**.

technique can identify differences in interactions between different proteins. For all measurements the Protein G background has been subtracted as shown in **Figure 6.2.C**. The background is probably due to the presence of Tyr in Protein G.

6.2 Trp-Cys interaction potential

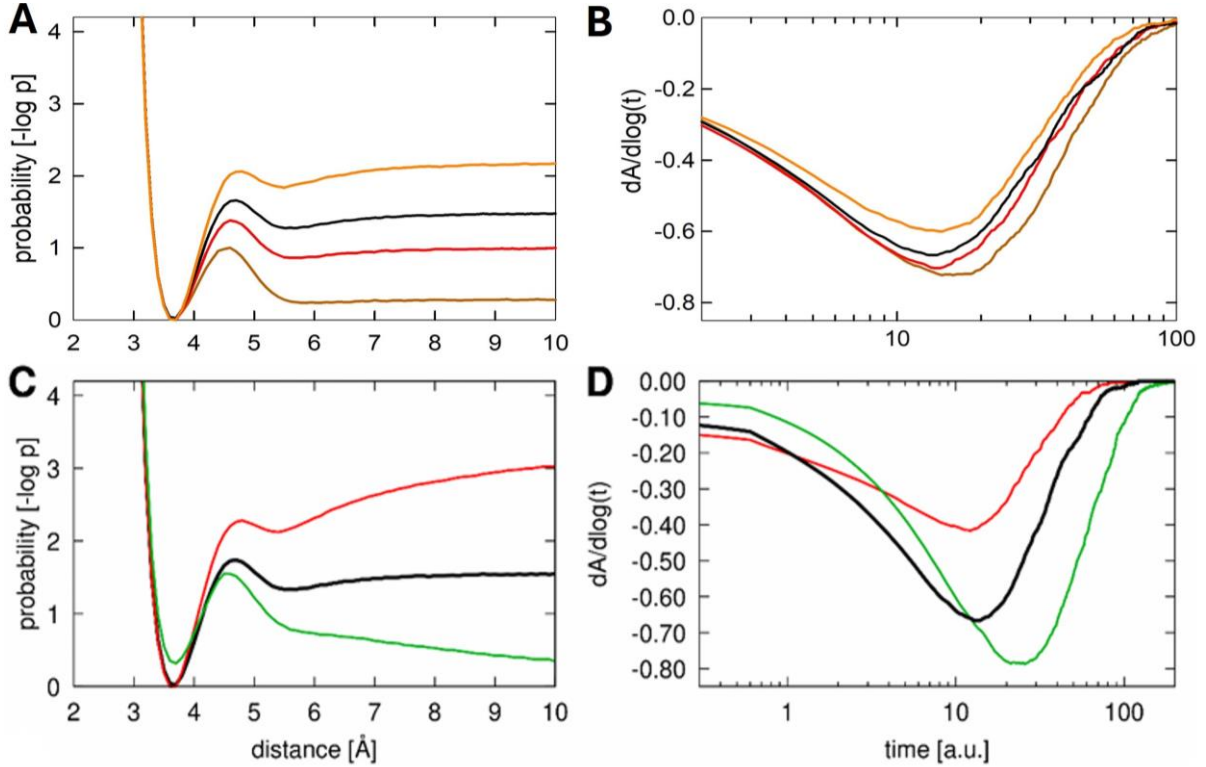


Figure 6.3: 1D potential analysis. Probability distribution functions are shown on the left (**A, C, E**), and derivatives of the calculated decay curves are shown on the right (**B, D, F**). The reference distribution in black is based on the Trp-Cys contact probability extracted from atomistic simulations for the Villin WT. Variations in the potential near the contact minimum are shown in **A** and **B**. Variations in long-range attraction ($d = -1$ in Eq. 6.1, red) or repulsion ($d = 1$ in Eq. 6.1, green) are shown in **C** and **D**.

To interpret the quenching curves, as a first approach we modeled the Trp-Cys quenching distance for Protein G with Villin WT using Monte Carlo (MC) simulations on a 1D potential as indicated by Eq. 6.1. The potential combines a Lennard-Jones type contact potential with a Gaussian potential to model a kinetic barrier and a long-range Debye-Hückel type potential and sampled at $kT = 1$. Trial moves were taken randomly in a positive or negative direction with a step size chosen randomly up to a maximum of 5Å .

$$V(r) = \varepsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right) + a e^{-(r-\mu)^2/w} + \frac{d \cdot k}{r} e^{-r/k} \quad (\text{Eq. 6.1})$$

The MC potential was approximated to the Trp-Cys contact probability obtained by atomistic (MD) molecular dynamics simulations (**Section 6.3**). The approximated distribution for Villin WT is shown in **Figure 6.3**. The potential parameters were further adjusted for Villin WT to impose variations in the potential depths at the contact minimum and the kinetic barrier height as shown in **Figure 6.3.A**. The calculated decay curves to shift slightly in time, but the derivative minimum varies significantly in depth (**Figure 6.3.B**). With a shallower contact minimum indicating weaker interactions we observe a deeper derivative value while a shallow derivative is observed for a deeper contact minimum indicating stronger interactions. The observed differences among the Villin variants may be the result of variations in potential at the contact minimum. While Villin WT and the double-mutant have contacts at residue 24 that show a shallower derivative, Villin V10W and K33W have contacts at residues 10 and 33 respectively, that show a deeper derivative (**Figure 6.3.B**). It is possible that Villin V10W and K33W weakly interact with Protein G compared to Villin WT and the double-mutant due to the location of the contact. In addition, the observed deeper minimum for Villin double-mutant vs WT is presumably due to its more negative charge resulting in weaker interactions.

To test the effect of long-range interactions on decay curves we adjusted d in **Eq 6.1** to 1 or -1 to introduce repulsion or attraction respectively (**Figure 6.3.C**). The quenching curves shifted to longer time scales with a deeper derivative minimum in the case of repulsion and to a shallower derivative minimum in the case of attraction with no shift in time (**Figure 6.3.D**). Repulsion between two proteins leads to less frequent contacts. Experimentally the decay of SH3 with Protein G shows a shift to longer time scales compared to the Villin variants at the same concentration indicating repulsion and less frequent contacts. This can be expected as both SH3 (-4) and Protein G (-5) have net negative charges, whereas Villin variants have a net positive or slightly negative (-1 to +2) charge.

6.3 Survival probability

To interpret the experimental observations at the molecular level, we applied μ s-scale MD simulations and analyzed the dynamics and interactions between the proteins in the experimentally measured systems. MD simulations were carried out using Anton2[56], OpenMM[46] and c36m[57] CHARMM force fields. One probe and nine quencher protein structures were randomly

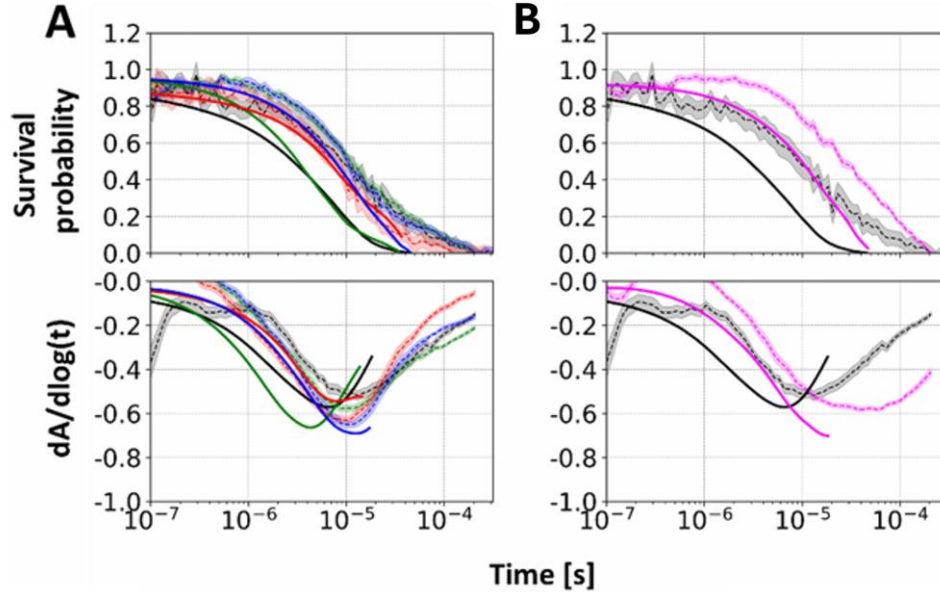


Figure 6.4: Comparing atomistic survival probabilities to experimental Trp decays. The probabilities against log-time are shown on the top panel and their derivatives in the bottom panel. Derivative values were obtained from a linear fit of probabilities against log-time over 21 data points. Experimental curves are shown as dashed lines with transparent shades for standard errors. Matching simulation curves are shown as solid lines. **A)** Results for Villin WT (black), V10W (red), K33W (blue), and the R15T+K30E (green). **C)** Results for Villin WT (black) and SH3 (magenta).

oriented and placed in a cubic box with a width of 144.048 Å, corresponding to concentrations of 0.56 and 5.0 mM for the probe and quencher proteins, respectively. The rest of the cubic box was filled with explicit water molecules and 25 mM of sodium and chloride ions to achieve an overall neutral system composed of around 297,000 atoms. A detailed description of the atomistic simulation procedure can be found in [1]. The concentration of the quencher proteins is 5-fold higher than was achieved in the experiment due to computational costs of simulating larger boxes and solubility issues with the K10C mutation of protein G. Therefore, we used coarse-grained simulations to rescale the survival probability to match the experimental conditions, as described in **Section A5**.

The survival probabilities of the triplet state of Trp were calculated as indicated in **Figure 6.4** from the simulation according to **Eq.6.2** based on Trp-Cys distance-time series sampling.

$$S_{quenching}(t) = \langle \exp\left(-\int_0^t q(t')dt'\right) \rangle = \frac{1}{t_{max}-t} \sum_0^{t_{max}-t} \exp\left(-\int_0^t q(t')dt'\right) \quad (Eq. 6.2)$$

Here $q(t')$ is the exponential quenching function as defined by **Eq 2.1**, where $a = 3.5$ Å. The triplet-to-singlet transition due to the natural decay of Trp was considered via **Eq. 6.3** with a rate constant of $k_0 = 2.3 \times 10^{-5} \text{ ns}^{-1}$ [35].

$$S(t) = S_{\text{quenching}}(t) \times e^{-k_0 t} \quad (\text{Eq. 6.3})$$

MD simulations (**Figure 6.4.A and B**) were able to capture many of the features observed in experiment. The decay derivative of Villin K33W shows excellent agreement with experiment. The double-mutant has a deeper derivative compared to Villin WT as in experiments indicating weaker interactions, but they do have a shift in time as well as downwards. Villin V10W shows good agreement in time, but it does show a shallower minimum compared to experiment. SH3 is in good agreement with experiment, capturing the shift to longer time scales and a deeper derivative compared to Villin WT.

Solid lines in **Figure 6.5** show the calculated distance probability distributions between Trp and Cys considering only the minimum contact distances. A contact minimum and a kinetic barrier were observed. The probabilities were fitted with the MC 1D potentials indicated by dashed lines for each protein system. Parameters are listed in **Table A4**. We see distinguishable differences in

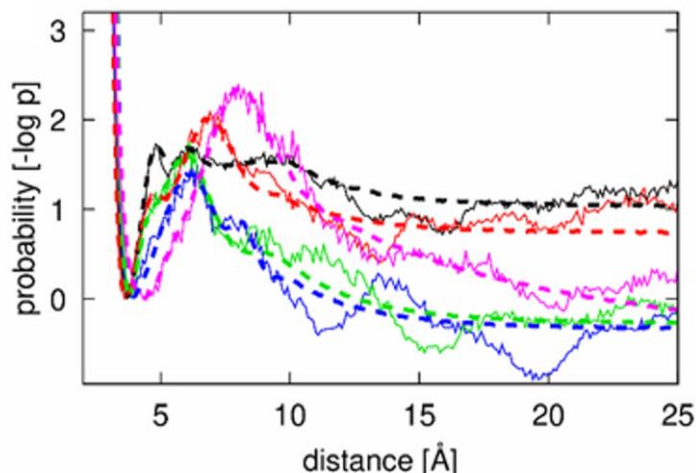


Figure 6.5: Interaction potentials for the protein systems. A) Probability distribution functions. Solid lines represent MD simulation results, and the dashed lines shows the 1D MC potential fits using **Eq 6.1**. Different colors reflect different probes: Villin WT (black), Villin V10W (red), Villin K33W (blue), Villin R15T+K30E (green) and SH3 (magenta).

the contact minimum, followed by a kinetic barrier and a decline of potential beyond 10 Å. This can be expected since Protein G may interact with the proteins in many arrangements that do not place protein G's C10 near the Villin/SH3 Trp. Above 15 Å the probabilities almost plateau for the Villin variants while the decline continues for SH3.

6.4 Surface preference for interactions

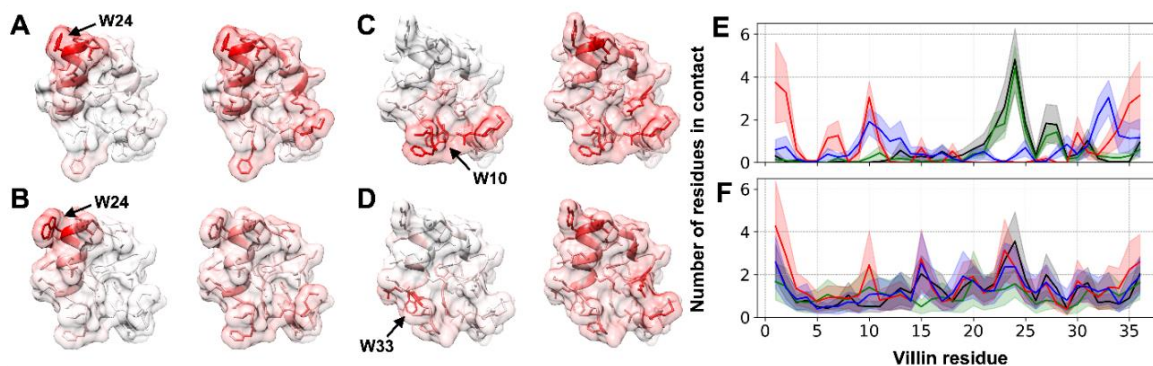


Figure 6.6: Residue-wise contacts between Protein G and Villin variants, projected onto the Villin surface (A–D) and as a function of Villin residue index (E, F). Surface projections are shown for Trp-Cys quenching contact (left) and for all contact positions (right) for **A)** Villin WT, **B)** Villin R15T+K30E, **C)** Villin V10W and **D)** Villin K33W. The location of the Trp residue is indicated by arrows. Contacts per frame vs residue index are shown **E)** at the time of quenching contact and **F)** at any time of contact for Villin WT (black), V10W (red), K33W (blue), and R15T+K30E (green). Shaded areas indicate standard errors. Contacts were defined by residue pairs whose interatomic distances were closer than 5 Å.

We then investigated the interactions between Protein G and Villin variants from MD simulations (**Figure 6.6**). We find that the overall interaction preference between Protein G and Villin are similar for all variants, and site 24 shows a slightly higher preference compared to elsewhere in the proteins. When we consider Protein G-Villin interactions at the time of Trp-Cys contact we clearly see that quenching near residue 24 is highly specific to the immediate vicinity of residue 24, whereas quenching near V10W and K33W is broader. This may be expected since the surface near 10 and 33 is flatter compared to 24 site pointing into the solvent at the tip of the Villin structure which can result in stronger interactions compared to sites 10 and 33. Comparing this observations to experimental data we see that, less frequently interacting V10W and K33W result in a deeper minimum while frequently interacting WT and the double mutant results in a shallower minimum (**Figure 6.2 B**) in agreement with the strength of short-range interactions (**Figure 6.3 A and B**).

Chapter 7

Dynamics and early-stage aggregation of O-GlcNAcylated α -synuclein

Here we study the effect of glycosylation on α -syn which is a major toxic protein in Parkinson's disease. Monomer dynamics and the early stage of aggregation were investigated for α -syn modified at two different sites T72 and S87. These two sites among others (T75, T81) have been shown to inhibit or delay α -syn fibril formation. While S87 slows down fibrilization, T72 exhibits heavy suppression. Therefore, these two were chosen to understand their distinct effects on the early stage of aggregation or small oligomer formation that takes place well before fibrilization of α -syn.

The α -syn used in this study is mutated at sites A69 with a Cys and F94 with a Trp to carry out Trp-Cys quenching experiments (Section 2.1). α -syn69c94w is modified with O-GlcNAc at sites T72 and S87 which are termed here as α -syn(gT72) and α -syn(gS87), respectively. Proteins were expressed by the Matt Pratt group at USC[58].

7.1 Fibrilization of O-GlcNAcylated α -synuclein

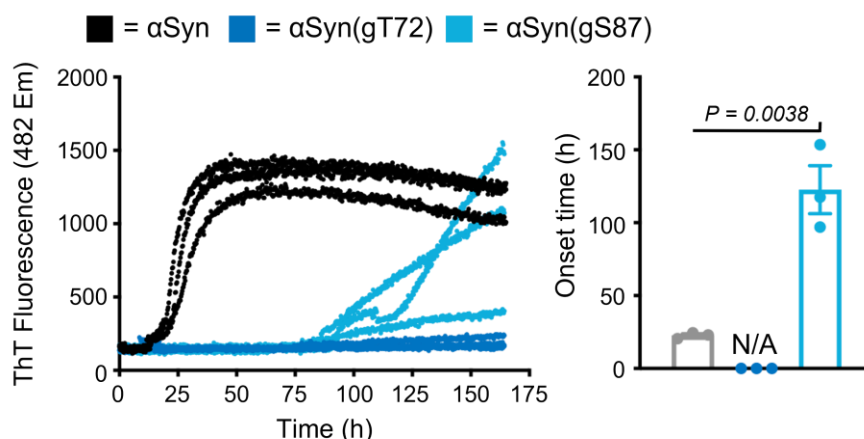


Figure 7.1: α -syn fibrilization kinetics. ThT fluorescence ($\lambda_{ex} = 450$ nm, $\lambda_{em} = 482$ nm) of α -syn and its glycosylates depicting fibrilization and delayed fibrilization respectively (Left). Three replicates are shown for each measurement. The fibrilization onset-times (Right) of the proteins represented as mean $\pm$ standard error of the mean. Statistical significance was determined using a two-way, unpaired Student's t test.

Thioflavin T (ThT) fluorescence measurements were carried out by the Pratt group for α -syn, α -syn(gT72) and α -syn(gS87) to analyze the fibrilization behavior of the proteins and to be certain that the 69C and 94W mutations don't affect fibrilization. ThT is a dye that can detect the presence

of amyloid fibrils. It attaches to stacked β -sheets of amyloid fibrils and exhibit enhanced fluorescence upon binding. 50 μ M of each protein in phosphate buffer saline (PBS) were agitated at 37°C in a plate reader to achieve aggregation. 10 μ M ThT was added to each of the solutions and fluorescence measurements were taken every 15 minutes over 7 days. α -syn69C94W exhibited the same fibrilization behavior as expected for wild type α -syn[59] indicating that the mutations do not significantly affect fibrillization kinetics of the protein. α -syn(gT72) showed delayed fibrilization while α -syn(gT72) almost showed none over the measured time period similar to the glycosylated wild type α -syn[59]. We also calculated the onset times for aggregation (2.5 times the initial ThT value) for α -syn and α -syn(gS87) and confirmed that the differences in aggregation kinetics were statistically significant.

7.2 Intra-molecular diffusion

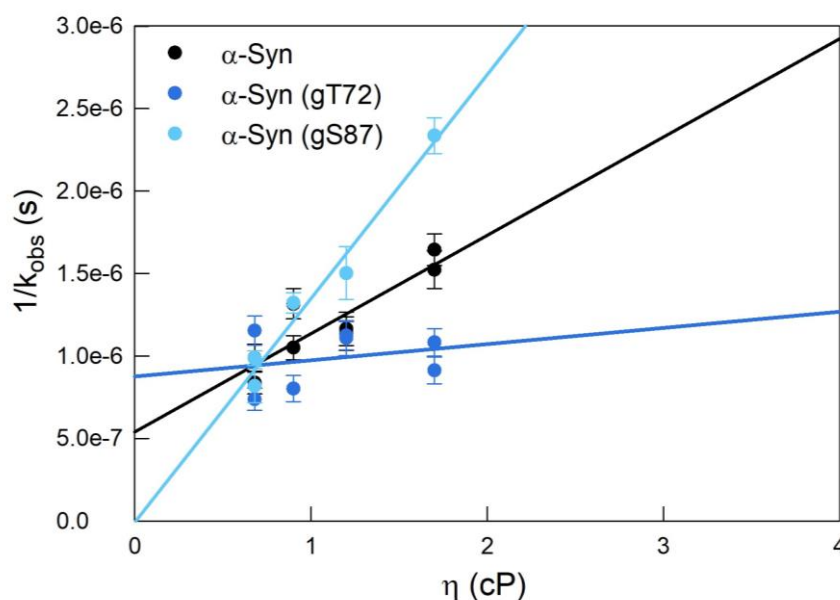


Figure 7.2: Measured decay rates as a function of viscosity: $1/k_{obs}$ vs η is plotted for α -syn (black), α -syn(gT72) (dark blue) and α -syn(gS87) (light blue) at 37°C. Data is shown for two sets. Decay rates and their standard errors were obtained from 1st order exponential fits.

Trp-Cys measurements were carried out using a pump-probe spectroscopic setup (Section 2.3) for the modified and unmodified α -syn at pH 7.4 and 37°C for different viscosities which was controlled by the addition of sucrose. Lyophilized powders of α -syn, α -syn(gT72) and α -syn(gS87) were dissolved in 20 mM sodium phosphate buffer, sonicated for 15 minutes and the solutions were centrifuged at 12000 rpm for 2 minutes to get rid of the insoluble fractions. 20 mM sodium

phosphate buffer with and without 50 % w/v sucrose were bubbled separately with N₂O for an hour in sealed vials to deoxygenate and scavenge free electrons (**Section 2.3**). To minimize the sample usage a syringe pump system (**Section 2.4**) was used to inject the proteins at final concentrations of 20 μ M, 16 μ M and 12 μ M of α syn, α -syn(gT72) and α -syn(gS87) respectively into the cuvette together with 0 % and 50 % w/v sucrose buffers to get final viscosities at 0, 10, 20 and 30 % w/v.

Figure 7.2 shows the Trp decay plotted against the viscosity for each protein. The intercept of each fit is equal to $1/k_R$ and the slope of each fit is equal to $1/\eta k_{D+}$. For α -syn(gS87) the intercept ($1/k_R = -6.9 \times 10^{-9} \pm 1.5 \times 10^{-7}$ s) is consistent with zero, therefore k_R cannot be determined but the lower limit of this rate can be estimated from the error in the intercept. Similarly, the slope ($1/\eta k_{D+} = 9.8 \times 10^{-8} \pm 1.7 \times 10^{-7}$ s $\cdot$ cP⁻¹) of α -syn(gT72) is also consistent with zero. Therefore, we assign the inverse of the upper limit of the 95 % confidence interval to k_R of α -syn(gS87) and k_{D+} of α -syn(gT72) respectively. k_{D+} is calculated at the viscosity of water, $\eta = 0.68$ cP. The diffusion-limited and the reaction-limited rates were extracted from the linear fits and plotted in **Figure 7.3.A and B** respectively. To calculate the intra-molecular diffusion coefficient (D) from these measured rates we follow the SSS theory (**Section 2.2**). We choose $a = 4$ Å which is the van der Waals distance between a Trp and a Cys and approximate a Gaussian chain model with a Trp-Cys distance probability distribution of

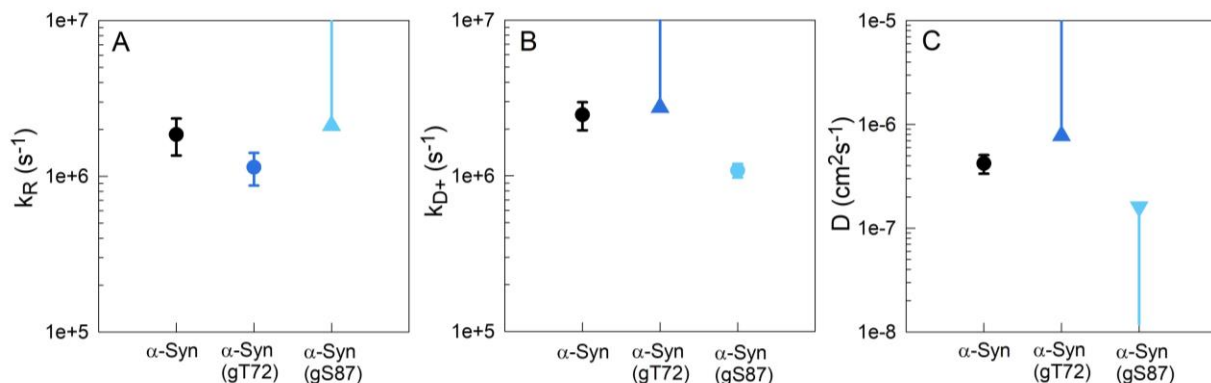


Figure 7.3: Computed rates and diffusion coefficients: For α -syn, α -syn(gT72) and α -syn(gS87) at 37°C. **A)** Reaction-limited rates. **B)** Diffusion-limited rates calculated for $\eta = 0.68$ cP. **C)** Diffusion coefficients. Triangles indicate the lower or upper limits computed in cases where the $1/k_R$ and $1/\eta k_{D+}$ values were consistent with zero.

$$P(r) = 4\pi r^2 \left[\frac{3}{2\pi \langle r^2 \rangle} \right]^{\frac{3}{2}} \exp \left[-\frac{3r^2}{2\langle r^2 \rangle} \right] \quad (\text{Eq. 7.1})$$

where $\langle r^2 \rangle$ is the mean squared contact distance which is adjusted in **Eq. 2.8** to match the measured k_R . **Figure 7.3.C** shows the calculated D . Due to the measurement limitations described above we estimate a lower limit for α -syn(gT72) and an upper limit for α -syn(gS87). Overall, we find $D_{\alpha\text{-syn(gT72)}} > D_{\alpha\text{-syn}} > D_{\alpha\text{-syn(gS87)}}$ (**Table 7.1**).

Table 7.1. Computed rates and diffusion coefficients of α -syn, α -syn(gT72) and α -syn(gS87) at 37°C. LL and UL represent the lower limit and the upper limit of the 95 % confidence interval of the measurements, respectively.

	k_R (s^{-1})	k_{D+} (s^{-1})	D (cm^2s^{-1})
α-syn	1853891.85 ± 494561.52	2466415.38 ± 502962.75	$4.21 \pm 0.86 \times 10^{-7}$
α-syn(gT72)	1142347.16 ± 270348.86	(LL) 2747859.19	(LL) 7.78×10^{-7}
α-syn(gS87)	(LL) 2118096.82	1084261.12 ± 109290.91	(UL) 1.61×10^{-7}

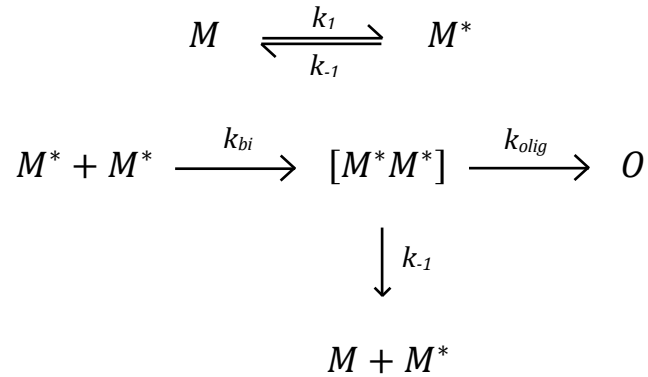
The intra-molecular diffusion of α -syn(gT72) is faster than the unmodified α -syn as well as α -syn(gS87) indicating its lower propensity to aggregate. In accordance with this we see that fibrilization of α -syn(gT72) is not observed even after several days as observed by ThT fluorescence. In contrast, α -syn(gS87) exhibits slower diffusion making it more prone to aggregation even though fibrilization is delayed compared to unmodified α -syn. This behavior can be true since ThT doesn't capture the formation of small oligomers well before fibrilization, therefore α -syn(gS87) can still form small oligomers slowing down its dynamics while delaying the oligomer-to-fibril transition. To understand this, we looked into the aggregation kinetics of the proteins.

7.3 The kinetic model of aggregation

To understand the early stage of aggregation of the proteins we hypothesize that when a protein has a faster reconfiguration rate it is more likely to escape from interacting with another slowing down the aggregation process while slow reconfiguration can provide ample time for the formation of bimolecular interactions and speed up aggregation. This basic model explains the relative fibrilization kinetics in **Figure. 7.1** for α -syn and α -syn(gT72), but not for α -syn(gS87).

To further understand how the measured diffusion coefficients can plausibly explain the observed fibrilization, we construct a kinetic model of early aggregation (Scheme 1) and fibrilization (Scheme 2). Scheme 1 assumes monomeric α -synuclein divided into two populations of non-aggregation-prone (M) and aggregation-prone (M^*) conformations, where the rate of formation of M^* by M and vice-versa are given by k_1 and k_{-1} respectively. The aggregation-prone population is more likely to have more solvent exposed hydrophobic regions than the non-aggregation-prone population. When two aggregation-prone conformations meet due to bimolecular diffusion at a rate of k_{bi} , they can form an encounter complex that can either form stabilizing interactions in an oligomer (O) at a rate of k_{olig} or disassociate due to one monomer changing its conformation to M at a rate k_{-1} .

Scheme 1:



To estimate reconfiguration and bi-molecular rates we use the relationship for the rate of diffusion of a particle under Brownian motion

$$k = \frac{2nD}{\langle x^2 \rangle} \quad (Eq. 7.2)$$

where D is the diffusion coefficient, x is the distance a particle diffuses, and n is dimension. We estimated the reconfiguration rates or the rates of change in conformation k_1 and k_{-1} as the time taken for a residue to diffuse across the diameter of the chain with intra-molecular diffusion coefficient as calculated experimentally (**Figure 7.3.C**). Therefore,

$$k_1 = k_{-1} = \frac{6D}{(2R_G)^2} \quad (Eq. 7.3)$$

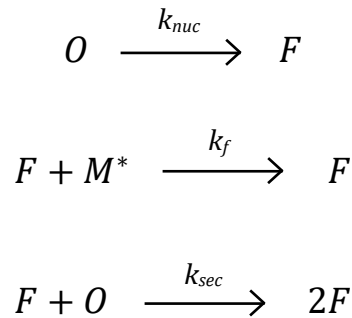
where $n = 3$ and $R_G \sim 3$ nm[39] is an estimated radius of gyration which should not change significantly with post-translational modification (PTM). This gives us values for k_1 as 7.0×10^6 s⁻¹, 1.3×10^7 s⁻¹ and 2.7×10^6 s⁻¹ for α -syn, α -syn(gT72) and α -syn(gS87) respectively. We then estimated k_{bi} based on **Eq. 7.2** as

$$k_{bi} = \frac{6D_{bi}}{r^2} \quad (Eq. 7.4)$$

which gave a value of 5.8×10^5 s⁻¹ where r is the distance between two closest molecules. $r = 3.2 \times 10^{-6}$ cm (320 Å) was calculated by considering the molecules as point particles homogeneously spread out at a concentration of 50 μ M in solution to match the experimental conditions of fibrillization (**Figure 7.1**). We use $D_{bi} = 1.0 \times 10^{-6}$ cm²s⁻¹[35] as a typical inter-molecular diffusion coefficient for a 14 kDa protein. Solving this model for an arbitrary rate of $k_{olig} = 10$ s⁻¹ the oligomer populations are plotted against time in **Figure 7.4.A**. α -syn(gS87) has the fastest formation of oligomers and α -syn(gT72) has the slowest formation proportionate with the differences in k_1 .

Subsequently we consider fibril formation according to scheme 2 where primary and secondary nucleation and elongation of fibrils are also considered, along with processes in Scheme 1.

Scheme 2:



Using arbitrary rates of $k_{nuc} = 0.0001 \text{ s}^{-1}$, $k_{sec} = 10 \text{ s}^{-1}$ and $k_f = 10 \text{ s}^{-1}$ we see the early-stage oligomer formation arrests (**Figure 7.4.B**) near the onset of fibril formation (**Figure 7.4.D**) and the oligomerization and fibrilization agree with each other for all the three proteins. While oligomer formation is fastest for α -syn(gS87) so is the fibril formation and at the same time α -syn(gT72) is the slowest in oligomer and fibril formation. But these results don't agree with experimentally observed ThT measurements (**Figure 7.1**). By reducing k_f to 0.1 s^{-1} for the modified proteins we get good qualitative agreement with experiment where fibrilization is delayed compared to unmodified α -syn (**Figure 7.4.E**) while the oligomer formation is dominant in the order of α -syn(gS87) > α -syn > α -syn(gT72) (**Figure 7.4.C**).

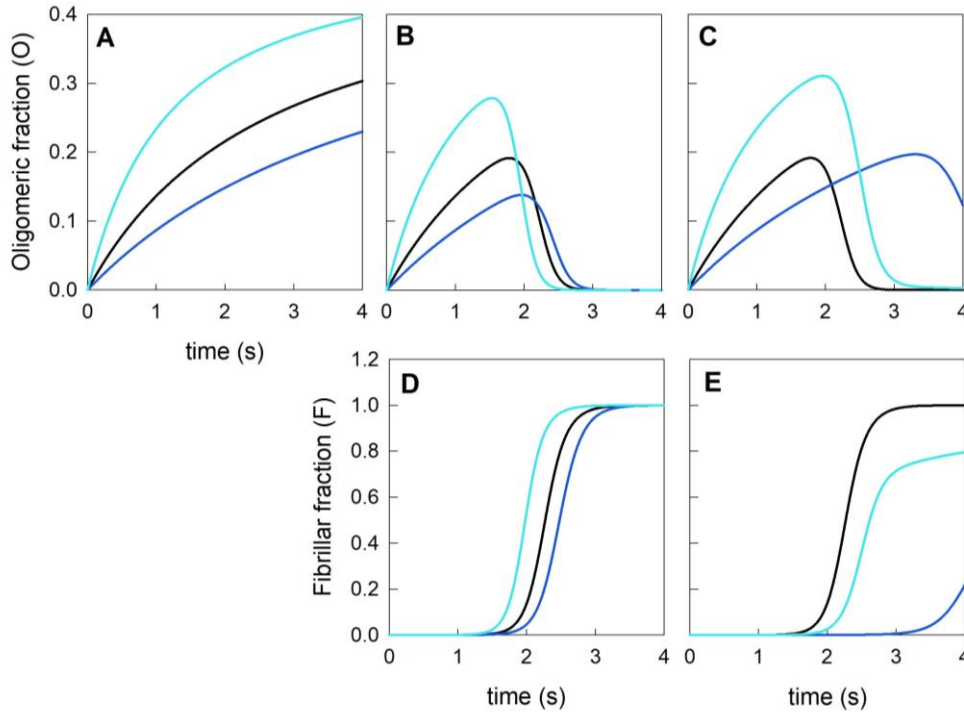


Figure 7.4: Kinetic model of aggregation: **A)** Formation of oligomers using Scheme 1 for $k_I = k_{-I} = 4.7 \times 10^6 \text{ s}^{-1}$ (black, unmodified protein), $k_I = k_{-I} = 7.8 \times 10^6 \text{ s}^{-1}$ (dark blue, α -syn(gT72)) and $k_I = k_{-I} = 1.6 \times 10^6 \text{ s}^{-1}$ (light blue, α -syn(gS87)). $k_{bi} = 9.7 \times 10^4 \text{ s}^{-1}$ and $k_{olig} = 10 \text{ s}^{-1}$ are the same for all proteins. **B)** Formation of oligomers and **D)** formation of fibrils using Scheme 1 and Scheme 2. Rates are the same as for **A** with the addition of $k_{nuc} = 0.0001 \text{ s}^{-1}$, $k_f = 10 \text{ s}^{-1}$, $k_{sec} = 10 \text{ s}^{-1}$. **C)** Formation of oligomers and **E)** formation of fibrils. The rates are same as for **B** and **D** except $k_f = 0.1 \text{ s}^{-1}$ for α -syn(gT72) and α -syn(gS87).

The delay in fibrilization for the modified mutants can be due to steric hindrance arising from the addition of the glycosyl group. Previous studies have shown that the addition of the glycosyl at S87 can alter the structure of α -syn and can inhibit its aggregation seeded by unmodified α -syn fibrils[60]. It is also observed that gS87 is less inhibitory than gT72 in agreement with the observed

fibrilization data. But this cannot be explained by steric hindrance, therefore it may be entirely due to variations in intra-molecular diffusion and consequent oligomer formation as observed by the kinetic model.

According to the kinetic model we see that PTM can either slow down or speed up oligomerization depending on the site. The reason might be that the most likely intra-molecular interactions for these two sites are different. Studies show that S87 makes preferential contacts at close proximity on the C-terminal side (sites A90, 90-100) while T72 prefers to interact with sites far away in sequence on the N-terminal side (sites A18, S42, Q62)[61, 62]. This behavior is also observed from molecular dynamic (MD) simulations[36]. Therefore, we hypothesize that glycosylation at T72 disrupts transient interactions between residues near T72 and far away in the sequence facilitating interactions at close proximity and making the chain more diffusive, while glycosylation at S87 disrupts interactions at the vicinity of the site itself, allowing those sites to make preferential contacts further away in the sequence and making the chain less diffusive (**Figure 7.5**). This claim is also supported by previously observed faster diffusion rates of the mutant T72P[53].

Overall, we see inhibition of α -syn fibril formation with glycosyl modifications but gT72 is more effective than gS87. At the same time the inhibition effect is more distinguishable at the oligomer stage where gS87 is less diffusive suggesting an increase in toxicity of α -syn while T72 is more diffusive which can reduce the toxicity. Therefore, we observe that the toxicity of glycosylated α -syn can vary depending on the modified site.

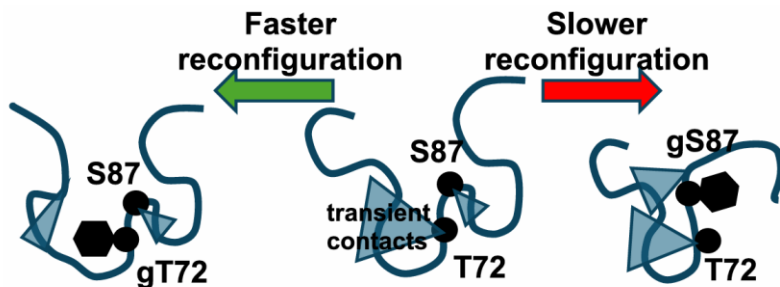


Figure 7.5: Site dependent reconfiguration. O-GlcNAcylation of α -syn69c94w at site T72 disrupts long-range transient interactions facilitating short-range interactions resulting in faster reconfiguration while site S87 disrupts short-range transient interactions facilitating long-range interactions slowing down reconfiguration.

Chapter 8

Conclusion

The cellular environment is crowded with high concentrations of proteins, nucleic acids and other macromolecules where a species can interact in a multitude of ways. This complex environment is not easily mimicked *in vitro*. Most of the time crowding is achieved through synthetic polymers such as PEG, Ficoll, dextran and sucrose or concentrating the protein of interest itself which can underestimate interactions and hide the true protein behavior and limiting the number of macromolecular systems that can be studied under physiologically relevant crowded conditions. We find condensates formed through LLPS to be a better platform to study proteins or RNA and this work investigated the LLPS behavior and introduced techniques to analyze protein dynamics and interactions in condensate or crowded systems.

From the experiments carried out on the PolyA-RGRGG system we were able to show that short peptides and RNA of opposite charges can undergo length dependent LLPS and form condensates even with very short polymers as small as 5 residues in length, making this one of the smallest LLPS polymer systems observed. However, the shortest peptide [RGRGG]₁ does not sufficiently interact to form condensates, but it does act as a driver of condensation for longer polymers by lowering the saturation concentration of phase separation. Analysis carried out on the thermodynamics of the system with the aid of CG simulations indicated that phase-separation is driven by enthalpy, probably due to electrostatic interactions between the positive charged peptide and the negative charged RNA. But the compositions and enthalpies in condensates are unchanged with length indicating confinement entropy as the main driving force of the length dependence of LLPS. This was observed to be a result of enthalpy scaling with the number of residues while the entropy scales with the number of polymers in the system. This work introduces a quantitative platform to investigate thermodynamics underling LLPS and to predict phase separating conditions through the COCOMO CG model. In addition, we found that even the shortest of polymers can contribute to LLPS as co-condensers (PolyA₅) or drivers ([RGRGG]₁).

We developed the Trp-Cys quenching technique to measure the intra-molecular diffusion coefficient of an IDP in condensates with the aid of molecular dynamic simulations and confocal microscopy. The self-condensing RLP proved to be well suited to incorporate IDPs into a crowded

environment and carry out bulk spectroscopy measurements at physiological conditions to analyze protein behavior. RLP was able to recruit 94 % of the protein of interest α -syn into the condensed phase resulting in a concentration as high as ~ 300 times relative to the dilute phase. Analyzing the dynamics of α -syn we noticed that it is highly dynamic in the condensed phase showing only ~ 50 % slowdown in intra-molecular diffusion compared to its monomeric state in the dilute solution even under these highly dense conditions. However, translational diffusion of α -syn was observed to be much slower. This indicates that even though the highly crowded environment in the condensates reduces the bulk movement, the intra-chain dynamics are not significantly affected or slowed down. Overall, interactions are observed to be electrostatically driven in the condensed phase. Similar condensate systems[4] have shown ~ 30 times slowdown in translational motion while only a ~ 3 times decrease in protein reconfiguration rates under crowded conditions of ~ 1000 and ~ 300 times the densities and viscosities of the dilute solution respectively. They claim the highly dynamic behavior to be a result of rapid switching of pico- to nanosecond transient contacts.

Future work will be directed towards analyzing the dynamics of α -syn in the condensed phase using RLP as a scaffold or PEG as a crowder using FCS as an alternate method. This single-molecular level technique will provide more detailed dynamics of proteins in the condensed phase such as absolute translational diffusion and possible variations in intra-molecular diffusion. Protein properties in condensates such as absolute densities and concentrations will also be studied via FCS and confocal microscopy. The work carried out here results in the advancement of the Trp-Cys quenching technique towards more complex macromolecular systems. Furthermore, the dynamics of α -syn measured here together with densities can be convolved with atomistic simulations to get a detailed understanding of how the protein interacts in condensates. Preliminary data from such simulations has suggested very good agreement with the experimental results.

Using a folded protein system of Villin variants or drkN SH3 domain with the crowder Protein G we have advanced the Trp-Cys quenching technique to probe transient protein-protein interactions under crowded conditions. Characterizing transient interactions between a pair of proteins in a crowded system via experiment is difficult since these interactions are short-lived, don't exhibit strong preferences for specific contacts and vary at very short contact distances. This is where the

Trp-Cys quenching technique comes to aid. Its sub-nanometer length resolution and nano-scale time resolution make it possible to distinguish variations at very short distances and time scales. Our results demonstrate that the Trp-Cys quenching technique can capture differences between long-range repulsive and attractive interactions of proteins, sensitive to the strength of short-range interactions, identifies concentration variations and can recognize interaction preference for different locations. Moving the Trp and the Cys to different locations on the proteins will provide a complete map of interaction preferences. We expect that the advances in the technique achieved here can be convolved with the advances in **Chapter 5** to investigate inter-molecular dynamics of proteins in LLPS systems.

Post-translational modification of α -syn by O-GlcNAc is known to inhibit fibrilization regardless of the modified site. But, based on the calculations of intra-molecular diffusion of O-GlcNAcylated α -syn done here, we have shown that the propensity to form early-stage oligomers depends on the glycosylated site. Glycosylation at site T72 makes the protein less compact and more diffusive, suppressing oligomerization while S87 makes the protein more compact and less diffusive promoting oligomerization. Based on a kinetic model of earliest aggregation we were able to show that this discrepancy for S87 can be true. Further, we hypothesized that this distinct diffusive behavior of the two sites is due to site-specific transient interactions. O-GlcNAcylation at site T72 can disrupt long-range transient interactions facilitating short-range interactions resulting in faster reconfiguration while site S87 can disrupt short-range transient interactions facilitating long-range interactions slowing down reconfiguration. These findings show that post-translational modifications do not affect aggregation uniformly. The effect is site-specific, which determines the rate of fibrilization or oligomerization.

Overall, our findings advance the knowledge of LLPS and techniques to investigate protein behavior and related properties in a heavily crowded environment. In addition, this work provides novel details on the dynamics and aggregation propensity of the Parkinson's disease causing IDP α -syn.

Bibliography

1. Heo, L., et al., *Characterizing Transient Protein-Protein Interactions by Trp-Cys Quenching and Computer Simulations*. J Phys Chem Lett, 2022. **13**(43): p. 10175-10182.
2. Jussupow, A., et al., *COCOMO2: A Coarse-Grained Model for Interacting Folded and Disordered Proteins*. J Chem Theory Comput, 2025. **21**(4): p. 2095-2107.
3. Acharya, S., et al., *Effects of Mutations on the Reconfiguration Rate of alpha-Synuclein*. J Phys Chem B, 2015. **119**(50): p. 15443-50.
4. Galvanetto, N., et al., *Extreme dynamics in a biomolecular condensate*. Nature, 2023. **619**(7971): p. 876-883.
5. Valdes-Garcia, G., et al., *The effect of polymer length in liquid-liquid phase separation*. Cell Reports Physical Science.
6. Lipiński, W.P., et al., *Biomolecular condensates can both accelerate and suppress aggregation of α -synuclein*. Science Advances, 2022. **8**(48): p. eabq6495.
7. Ahmed, R., et al., *Client-scaffold interactions suppress aggregation of a client protein in model condensates*. Proceedings of the National Academy of Sciences, 2025. **122**(33): p. e2508403122.
8. Rodríguez, L.C., N.N. Foressi, and M.S. Celej, *Modulation of α -synuclein phase separation by biomolecules*. Biochim Biophys Acta Proteins Proteom, 2023. **1871**(2): p. 140885.
9. Lafontaine, D.L.J., et al., *The nucleolus as a multiphase liquid condensate*. Nature Reviews Molecular Cell Biology, 2021. **22**(3): p. 165-182.
10. Fox, A.H., et al., *Paraspeckles: Where Long Noncoding RNA Meets Phase Separation*. Trends Biochem Sci, 2018. **43**(2): p. 124-135.
11. Molliex, A., et al., *Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization*. Cell, 2015. **163**(1): p. 123-33.
12. Brangwynne, C.P., et al., *Germline P granules are liquid droplets that localize by controlled dissolution/condensation*. Science, 2009. **324**(5935): p. 1729-32.
13. Das, T., et al., *Tunable metastability of condensates reconciles their dual roles in amyloid fibril formation*. bioRxiv, 2025.
14. Zhang, R., et al., *Rim4-seeded stress granules connect temperature sensing to meiotic regulation*. Nat Commun, 2025. **16**(1): p. 5566.
15. Hoffmann, C., et al., *Synapsin Condensates Recruit alpha-Synuclein*. J Mol Biol, 2021. **433**(12): p. 166961.
16. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry*. Nature Reviews Molecular Cell Biology, 2017. **18**(5): p. 285-298.
17. Flory, P.J., *Thermodynamics of High Polymer Solutions*. The Journal of Chemical Physics, 1941. **9**(8): p. 660-660.
18. Huggins, M.L., *Solutions of Long Chain Compounds*. The Journal of Chemical Physics, 1941. **9**(5): p. 440-440.

19. Alberti, S. and D. Dormann, *Liquid-Liquid Phase Separation in Disease*. Annu Rev Genet, 2019. **53**: p. 171-194.
20. Shin, Y. and C.P. Brangwynne, *Liquid phase condensation in cell physiology and disease*. Science, 2017. **357**(6357).
21. Banani, S.F., et al., *Compositional Control of Phase-Separated Cellular Bodies*. Cell, 2016. **166**(3): p. 651-663.
22. Ishov, A.M., et al., *PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1*. J Cell Biol, 1999. **147**(2): p. 221-34.
23. Decker, C.J. and R. Parker, *P-bodies and stress granules: possible roles in the control of translation and mRNA degradation*. Cold Spring Harb Perspect Biol, 2012. **4**(9): p. a012286.
24. Dzuricky, M., et al., *De novo engineering of intracellular condensates using artificial disordered proteins*. Nature Chemistry, 2020. **12**(9): p. 814-825.
25. Cakmak, F.P., et al., *Prebiotically-relevant low polyion multivalency can improve functionality of membraneless compartments*. Nature Communications, 2020. **11**(1): p. 5949.
26. Takamuku, M., et al., *Evolution of α -synuclein conformation ensemble toward amyloid fibril via liquid-liquid phase separation (LLPS) as investigated by dynamic nuclear polarization-enhanced solid-state MAS NMR*. Neurochem Int, 2022. **157**: p. 105345.
27. Elbaum-Garfinkle, S., et al., *The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics*. Proc Natl Acad Sci U S A, 2015. **112**(23): p. 7189-94.
28. Taylor, N.O., et al., *Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching*. Biophys J, 2019. **117**(7): p. 1285-1300.
29. Dai, Y., et al., *Programmable synthetic biomolecular condensates for cellular control*. Nat Chem Biol, 2023. **19**(4): p. 518-528.
30. Perego, E., et al., *Single-photon microscopy to study biomolecular condensates*. Nature Communications, 2023. **14**(1): p. 8224.
31. Galvanetto, N., et al., *Material properties of biomolecular condensates emerge from nanoscale dynamics*. Proc Natl Acad Sci U S A, 2025. **122**(23): p. e2424135122.
32. Joshi, A., et al., *Single-molecule FRET unmasks structural subpopulations and crucial molecular events during FUS low-complexity domain phase separation*. Nat Commun, 2023. **14**(1): p. 7331.
33. Wen, J., et al., *Conformational Expansion of Tau in Condensates Promotes Irreversible Aggregation*. J Am Chem Soc, 2021. **143**(33): p. 13056-13064.
34. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Dynamics of intramolecular contact formation in polypeptides: Distance dependence of quenching rates in a room-temperature glass*. Physical Review Letters, 2001. **87**(25): p. 8101.

35. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Measuring the rate of intramolecular contact formation in polypeptides*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(13): p. 7220-7225.
36. Woodard, J., et al., *Intramolecular Diffusion in α -Synuclein: It Depends on How You Measure It*. Biophysical Journal, 2018. **115**(7): p. 1190-1199.
37. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Measuring Dynamic Flexibility of the Coil State of a Helix-forming Peptide*. Journal of Molecular Biology, 2002. **319**(1): p. 19-25.
38. Waldauer, S.A., O. Bakajin, and L.J. Lapidus, *Extremely slow intramolecular diffusion in unfolded protein L*. Proceedings of the National Academy of Sciences, 2010. **107**(31): p. 13713-13717.
39. Ahmad, B., Y. Chen, and L.J. Lapidus, *Aggregation of α -synuclein is kinetically controlled by intramolecular diffusion*. Proceedings of the National Academy of Sciences, 2012. **109**(7): p. 2336-2341.
40. Koo, H.J., M.Y. Choi, and H. Im, *Aggregation-defective alpha-synuclein mutants inhibit the fibrillation of Parkinson's disease-linked alpha-synuclein variants*. Biochem Biophys Res Commun, 2009. **386**(1): p. 165-9.
41. Bent, D.V. and E. Hayon, *Excited-State Chemistry of Aromatic Amino-Acids and Related Peptides .3. Tryptophan*. Journal of the American Chemical Society, 1975. **97**(10): p. 2612-2619.
42. Li, Z., W.E. Lee, and W.C. Galley, *Distance dependence of the tryptophan-disulfide interaction at the triplet level from pulsed phosphorescence studies on a model system*. Biophysical Journal, 1989. **56**(2): p. 361-367.
43. Strambini, G.B. and E. Gabellieri, *Phosphorescence from Trp-48 in azurin: influence of copper(II), copper(I), silver(I), and cadmium(II) at the coordination site*. The Journal of Physical Chemistry, 1991. **95**(11): p. 4352-4356.
44. Szabo, A., K. Schulten, and Z. Schulten, *First passage time approach to diffusion controlled reactions*. J Chem Phys, 1980. **72**: p. 4350-4357.
45. Valdes-Garcia, G., et al., *Modeling Concentration-dependent Phase Separation Processes Involving Peptides and RNA via Residue-Based Coarse-Graining*. Journal of Chemical Theory and Computation, 2023. **19**(2): p. 669-678.
46. Eastman, P., et al., *OpenMM 7: Rapid development of high performance algorithms for molecular dynamics*. PLOS Computational Biology, 2017. **13**(7): p. e1005659.
47. Dutagaci, B., et al., *Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments*. eLife, 2021. **10**: p. e64004.
48. Das, S., et al., *Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins*. Proc Natl Acad Sci U S A, 2020. **117**(46): p. 28795-28805.
49. Laghmach, R., et al., *RNA chain length and stoichiometry govern surface tension and stability of protein-RNA condensates*. iScience, 2022. **25**(4): p. 104105.

50. Sanchez-Burgos, I., et al., *RNA length has a non-trivial effect in the stability of biomolecular condensates formed by RNA-binding proteins*. PLOS Computational Biology, 2022. **18**(2): p. e1009810.
51. Tejedor, A.R., et al., *RNA modulation of transport properties and stability in phase-separated condensates*. Biophysical Journal, 2021. **120**(23): p. 5169-5186.
52. Kayitmazer, A., *Thermodynamics of complex coacervation*. Adv Colloid Interface Sci., 2017. **239**: p. 169-177.
53. Acharya, S., et al., *Effects of Mutations on the Reconfiguration Rate of alpha-Synuclein*. Journal of Physical Chemistry B, 2015. **119**(50): p. 15443-15450.
54. Gonnelli, M. and G.B. Strambini, *Phosphorescence lifetime of tryptophan in proteins*. Biochemistry, 1995. **34**(42): p. 13847-57.
55. Soumpasis, D.M., *Theoretical analysis of fluorescence photobleaching recovery experiments*. Biophys J, 1983. **41**(1): p. 95-7.
56. Shaw, D.E., et al. *Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer*. in *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2014.
57. Huang, J., et al., *CHARMM36m: an improved force field for folded and intrinsically disordered proteins*. Nature Methods, 2017. **14**(1): p. 71-73.
58. Gamage, K., et al., *O-GlcNAc Modification of α -Synuclein Can Alter Monomer Dynamics to Control Aggregation Kinetics*. ACS Chemical Neuroscience, 2024. **15**(16): p. 3044-3052.
59. Lewis, Y.E., et al., *O-GlcNAcylation of alpha-Synuclein at Serine 87 Reduces Aggregation without Affecting Membrane Binding*. ACS Chem Biol, 2017. **12**(4): p. 1020-1027.
60. Levine, P.M., et al., *alpha-Synuclein O-GlcNAcylation alters aggregation and toxicity, revealing certain residues as potential inhibitors of Parkinson's disease*. Proc Natl Acad Sci U S A, 2019. **116**(5): p. 1511-1519.
61. Bertonecini, C.W., et al., *Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein*. Proc Natl Acad Sci U S A, 2005. **102**(5): p. 1430-5.
62. Allison, J.R., et al., *A relationship between the transient structure in the monomeric state and the aggregation propensities of α -synuclein and β -synuclein*. Biochemistry, 2014. **53**(46): p. 7170-83.
63. Howell, R.C., T. Proffen, and S.D. Conradson, *Pair distribution function and structure factor of spherical particles*. Physical Review B, 2006. **73**(9): p. 094107.
64. Garaizar, A., et al., *Expansion of Intrinsically Disordered Proteins Increases the Range of Stability of Liquid-Liquid Phase Separation*. Molecules, 2020. **25**(20): p. 4705.

Appendix

A1 Avoiding Cy3 induced phase separation

We observed that the Cy3 in the labeled RNA induces phase separation by itself above a certain concentration threshold. This is likely due to dye hydrophobicity and stacking interactions in oligomers. To determine the threshold, the concentration of PolyA₁₀-Cy3 was varied from 0-100 μ M in a mixture of PolyA₁₀ and [RGRGG]₂ each maintained at a concentration of 1mg/ml. At this concentration the mixture doesn't phase-separate on its own. Bright-field microscopy was carried out using an AmScope compound microscope was used equipped with a 10 $\times$ objective (NA 0.25) or a 20 $\times$ objective (NA 0.4). Samples were imaged on 25 $\times$ 75 $\times$ 1 mm Alkali Scientific microscope slides. Microscope slides were cleaned with tap water, de-ionized water and 70% ethanol sequentially before use. Results showed that no condensates formed up to 20 μ M, but condensates were observed at 30 μ M and above, indicating that the threshold lies between 20 and 30 μ M (**Figure A1**). Therefore, Cy3-labeled RNA was kept at a low concentration of 5 μ M compared to unlabeled RNA in all measurements to prevent it from inducing phase separation.

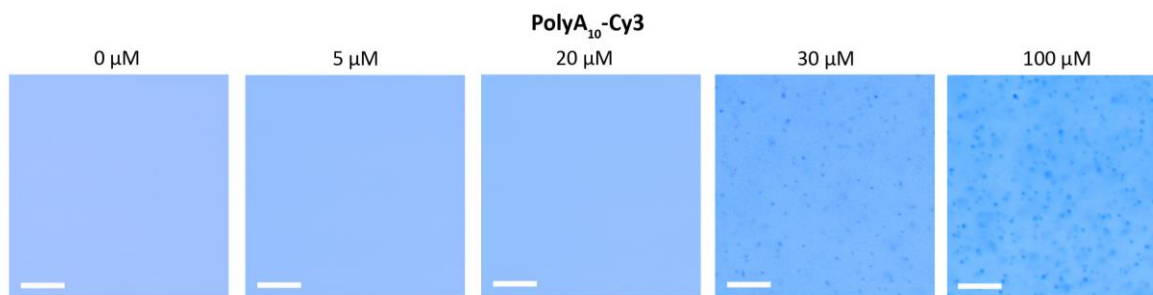


Figure A1: Inducement of condensates by Cy3. All samples contain PolyA₁₀ and [RGRGG]₂ at the same total concentrations of 1mg/mL with PolyA₁₀-Cy3 at 30 μ M and above showing condensates and no condensates were observed at and below 20 μ M. Scale bars represent 30 μ m.

A2 Free-energy model

A thermodynamic free-energy model based on enthalpy-entropy decomposition was created to explain the LLPS results we obtained through CG simulations and experiments. Phase separation requires the formation of the condensate to be energetically favorable relative to the initial disperse phase which is generally achieved through phase coexistence between the condensed and the dispersed phase that requires the chemical potential to be the same in both phases. But the present CG system doesn't account for phase coexistence therefore we focus here on just the free energy of the condensed phase relative to the disperse phase to determine under which conditions

condensate formation itself is favorable. In other words, this model focuses on estimating the stability of the condensates from the free energy of phase separation.

The total free energy of phase separation is estimated based on the free energies of condensed (ΔG_c), dilute (ΔG_u) and dispersed (ΔG_d) phases:

$$\Delta G_{total} = \Delta G_c + \Delta G_u - \Delta G_d \quad (Eq. A1)$$

Free energy is estimated from an enthalpy-entropy decomposition according to:

$$\Delta G = \Delta H - T\Delta S \quad (Eq. A2)$$

The enthalpy of a residue (i.e. arginine (R), glycine (G), or adenine (A)) in either phase was calculated as a sum of pair-wise interactions with other monomers in the same phase as well as self-interactions (interactions within a polymer):

$$\Delta h_{R,c} = \frac{1}{2}(\Delta h_{RR,c} + \Delta h_{RG,c} + \Delta h_{RA,c}) \quad (Eq. A3)$$

$$\Delta h_{G,c} = \frac{1}{2}(\Delta h_{GR,c} + \Delta h_{GG,c} + \Delta h_{GA,c}) \quad (Eq. A4)$$

$$\Delta h_{A,c} = \frac{1}{2}(\Delta h_{AR,c} + \Delta h_{AG,c} + \Delta h_{AA,c}) \quad (Eq. A5)$$

where the factor 1/2 corrects for double-counted interactions. Here ‘c’ denotes the condensed phase. Same equations are used in the dilute and the dispersed phases as well.

Each pair-wise enthalpy contribution was calculated from the convolution of pair-wise RDFs ($g(r)$) (**Figure 3.10**) with the interaction potential ($U(r)$) according to the coarse-grained interaction potential (**Eq. 2.10**):

$$\Delta h_{XY,c} = \rho_{X,c} \int_V \hat{g}_{XY,d}(r) U_{XY}(r) d^3r = 4\pi\rho_{X,c} \int_0^{r_{max}} \hat{g}_{XY,d}(r) U_{XY}(r) r^2 dr \quad (Eq. A6)$$

‘ X ’ and ‘ Y ’ denote the residue types. $\rho = \frac{n}{V}$ is the residue density in a given phase where ‘ n ’ is the number of residues and ‘ V ’ is the volume of the phase and ‘ r ’ is the radius determined as the distance from the center of the condensate to where the density drops to half of the density at the center of the condensate (**Figure 3.3.A**). The upper integration limit r_{max} was set to 15 nm for all interactions. At that radius and above, the interaction potential $U(r)$ becomes negligible. This integral form approximates the trajectory-averaged explicit summation of discrete particle interactions.

Because of the finite-size of the condensates, the extracted $g(r)$ functions for the condensed phase decrease with increasing radius according to[63]:

$$g(r) = \left[1 - \frac{3}{2} \left(\frac{r}{d} \right) + \frac{1}{2} \left(\frac{r}{d} \right)^3 \right] g_{\infty}(r) \quad (Eq. A7)$$

when determined from particles distributed homogeneously in a finite size sphere of radius d .

To compare finite size condensates with theoretically infinitely large condensates, we considered both the original $g(r)$ functions extracted from the simulations as well as normalized functions according to **Eq. A7** to obtain $g(r_{cut})=1$ at a given cutoff distance of r_{cut} (set to 12 nm) and a constant value of 1 for all larger radii. Such normalized $g(r)$ functions could then be rescaled again via Eq. S8 to reflect theoretical distribution functions for larger condensates such as those observed experimentally.

The enthalpy for a given phase was estimated by multiplying the enthalpies per residue according to **Eq. A3-A5** by the number of residues that were present in that phase. For example, the total enthalpy of the condensate was calculated according to:

$$\Delta H_c = n_{R,c} \cdot \Delta h_{R,c} + n_{G,c} \cdot \Delta h_{G,c} + n_{A,c} \cdot \Delta h_{A,c} \quad (Eq. A8)$$

The average numbers of residues in different phases for different systems were extracted from the simulations.

Enthalpies were calculated for the condensed and dispersed phases. Because of much lower monomer densities in the dilute phase, enthalpies of the dilute phase were less than 1 % of the condensate enthalpies for all systems considered and were neglected. The calculation of the total enthalpy of phase separation then becomes:

$$\Delta H_{total} = \Delta H_c - \Delta H_d \quad (Eq. A9)$$

To estimate the entropic cost of confinement of a polymer in the condensate, molar entropies were calculated from the ratio of available volume in the condensed phase to the system volume:

$$\Delta S_{X,c} = R \ln \left(\frac{V_{X,c}}{V} \right) \quad (Eq. A10)$$

where R is the universal gas constant and ‘X’ is either peptides or RNA. An important point here is that while enthalpy depends on residue-wise interactions, and therefore scales with residue densities, the loss of translational entropy applies to each polymer therefore scales with the number of polymers.

The total entropy due to confinement was then calculated by multiplying the molar entropies with the number of polymers in the condensate:

$$\Delta S_c = n_{P,c} \cdot \Delta S_{P,c} + n_{N,c} \cdot \Delta S_{N,c} \quad (Eq. A11)$$

where ‘P’ denotes peptide, ‘N’ denotes RNA and ‘n’ denotes the number of peptide or RNA.

We also considered a change in mixing entropy as the peptide/RNA ratio in the dispersed, dilute, and condensed phases may be different.

The mixing entropy for a given phase was calculated according to:

$$\Delta S_{mix} = (n_N + n_P) * (x_N \ln (x_N) + x_P \ln (x_P)) \quad (Eq. A12)$$

where x_p and x_N are the mole fractions of RNA and peptide, respectively, in the phase.

The entropic contribution to phase separation due to a change in mixing entropy is then calculated from the difference in mixing entropies between condensate and dilute phases vs. the dispersed phase:

$$\Delta S_{mix,total} = \Delta S_{mix,c} + \Delta S_{mix,u} - \Delta S_{mix,d} \quad (Eq. A13)$$

Because the change in mixing entropy was small (**Figure 3.14**), it was neglected.

This leaves the following contributions to be considered in estimating the total free energy of phase separation:

$$\Delta G_{total} = \Delta H_c - \Delta H_d - T\Delta S_c \quad (Eq. A14)$$

A3 Conformational, mixing and counterion entropy

To understand the change in conformational entropy during phase separation we examined radii of gyration (R_g) for both peptide and RNA in condensates and the dispersed phase. R_g remained the same after phase separation except for very long lengths (**Figure A2**). We observed that the very long RNA stays extended in the condensates relative to the dispersed phase. This type of extended conformations in the condensed phase have been observed with disordered proteins where steric hindrance is reduced and intermolecular contacts are maximized, thereby enhancing phase separation[64]. Additionally, we calculated the probability distance distribution $P(r)$ between the 1st and the 5th residue of peptide and RNA chains (**Figure A2**). They too do not show any difference before and after phase separation indicating that polymer conformation doesn't contribute to the entropy of phase separation in this system.

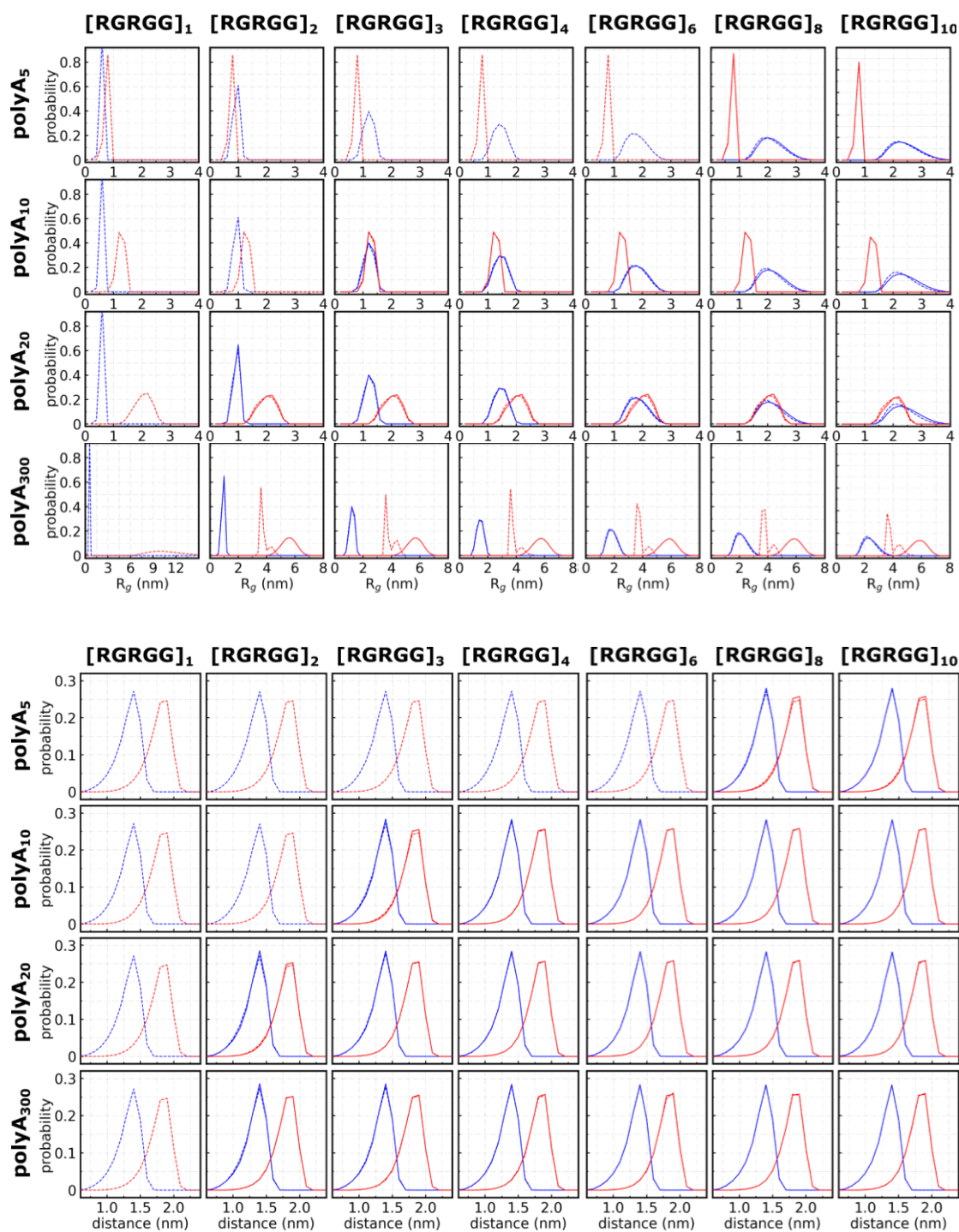


Figure A2: Radii of gyration and distance distributions: The histograms of radii of gyration (upper panels) and residue 1-5 probability distance distributions (lower panels) of peptide (blue) and RNA (red) in the condensed (solid lines) and dispersed (dashed lines) phase is shown here. Results are shown row-wise for PolyA_N (N = 5, 10, 20 and 300) and column-wise for [RGRGG]_M (M = 1, 2, 3, 4, 6, 8, and 10).

Mixing entropy is defined here as the entropy change due to the change in peptide to RNA molecular ratios when phase separating. The mixing entropies were calculated as shown in **Section A2**. The calculated $-T\Delta S$ values are less than 0.2 kJ/mol at 300 K for all mixtures. Therefore, mixing entropy can be neglected compared to the energy contribution from enthalpy and confinement entropy.

During the formation of condensates counterions associated with polymers can be released contributing to the entropy change of the system. This phenomenon is not explicitly parameterized in COCOMO. The CG simulations only account for charge shielding effects through the electrostatic term in the potential (**Table 2.2**). To examine the counterion effect the Feig group carried out explicit solvent all-atom simulations of peptides and RNA in the presence of 20 mM sodium phosphate at pH 8.0 and near experimental polymer concentrations of 1 mg/mL for peptides and 0.7 mg/mL for RNA. Results showed that only ~ 1 Na^+ ion associates with an RNA chain while ~ 7 phosphates associate with a peptide chain therefore the major contribution is from phosphate ions. Results also showed that phosphate ion association is not significantly changed with peptide length. Therefore, we conclude that the length dependence of phase separation is not affected by counterion entropy, but it most probably plays a role during phase separation[5].

A4 Trp decay transition from dilute to condensed phase

The Trp decay rate of α -syn39w69c was measured at several concentrations of RLP from 0.1 – 15 μM at 22°C in the presence of 18.33 mM NaCl as shown in **Figure A3**. At low concentrations of RLP below C_{sat} the dynamics of α -syn are faster and reach the decay rate of the dilute phase. The dilute phase rate is indicated by the cyan band. With the increase of RLP concentration (Above C_{sat} of RLP which is 0.86 μM) the samples are cloudier, and the OD of the samples increases gradually (**Figure A3**) indicating the increase in the number of condensates. With the formation of more and more condensates the decay rate captures the slower dynamics of α -syn in the condensed phase more than the faster dynamics in the dilute phase indicating a gradual increase in the fraction of α -syn recruited into the condensed phase. It was also noticed that the turbidity of RLP doesn't change with the addition of α -syn.

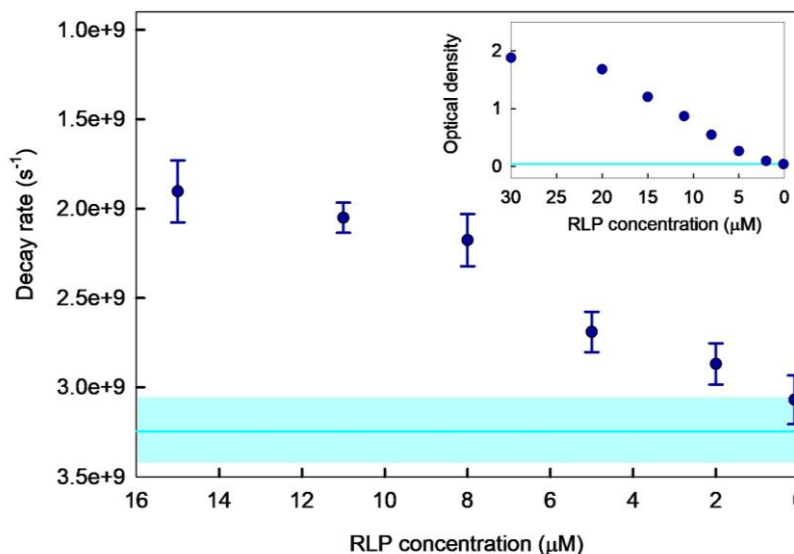


Figure A3: Trp decay transition from dilute to condensed phase. Decay rate is shown as a function of total RLP concentration. Trp decay rate shifts to slower dynamics above $C_{sat} = 0.86 \mu\text{M}$ while it becomes faster and reaches the dilute phase rate below C_{sat} . Measurements were taken with $\alpha\text{-syn}^{39\text{w}69\text{c}}$ at 22°C in the presence of 18.33 mM NaCl . The buffer used is $50 \text{ mM Tris pH } 7.4$. $\alpha\text{-syn}$ was kept at $50 \mu\text{M}$ while RLP was varied from $0.1 - 15 \mu\text{M}$. The cyan thin and thick lines represent the magnitude and the error of the $\alpha\text{-syn}$ decay rate in the dilute phase respectively. **Inset: OD at different RLP concentrations.** OD was calculated for RLP samples with a 1 cm path length. A higher OD is observed at higher RLP concentrations due to the presence of larger number of condensates. The cyan line indicates the OD of a sample without RLP.

A5 Concentration dependence

Experimental results clearly indicate a decay shift to faster times at higher crowder concentrations (Figure 6.2.A). We carried out CG simulations using COCOMO CG model (Section 2.5) to understand this effect. Simulations were performed for $60 \mu\text{s}$ with an time step of 20 fs , and snapshots were saved every 100 ps . The Debye-Hückel screening length was set to $\kappa = 12 \text{ \AA}$ to match the ion concentration in atomistic simulations of 25 mM . To maintain folded structures in proteins an additional distance restraint was applied with a force constant of $2.092 \text{ kJ/mol/\AA}^2$ to $\text{C}\alpha$ atom pairs that were separated by two or more residues and that had distances below $20 \text{ \AA}$.

We assumed that the concentration of proteins would not alter the shape of survival probability curves significantly but only shift them in time. Survival probability curves at various protein concentrations were evaluated using Eq. 6.2 and 6.3. Figure A4 shows results for the Villin and SH3 at $50 \mu\text{M}$ with Protein G at various concentrations of $0.5, 0.8, 2$ and 5 mM indicating a shift to faster times similar to experiment. Simulations were performed with one probe together with

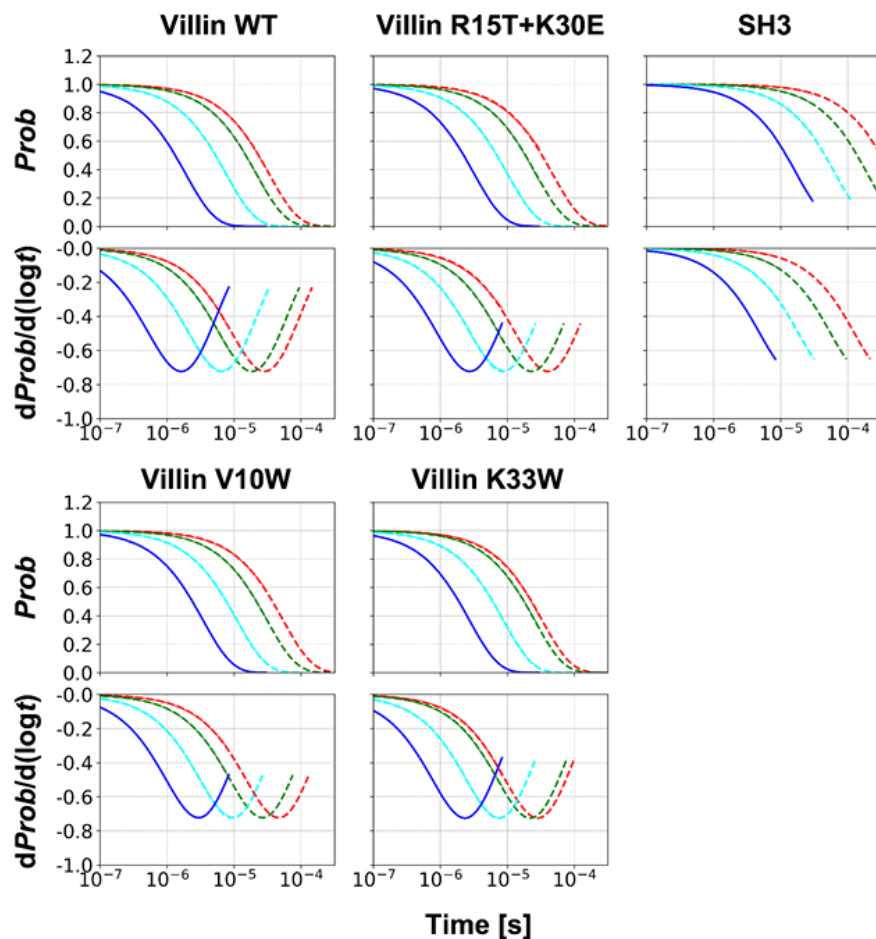
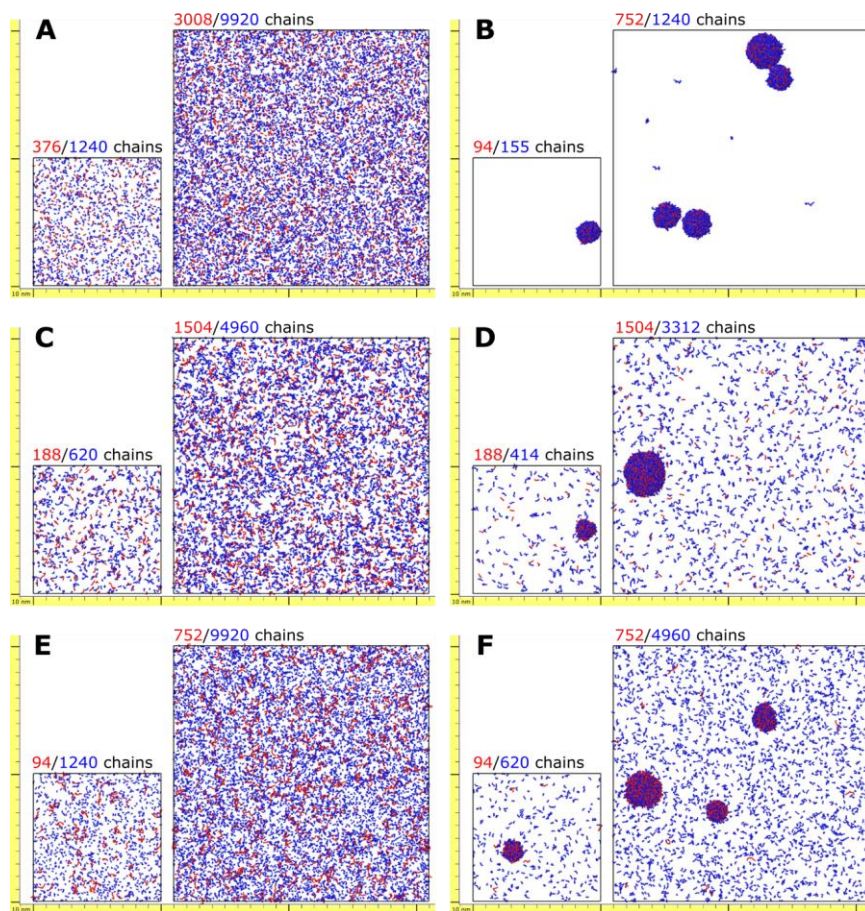


Figure A4: Survival probability as a function of quencher concentration from CG simulations. Simulations at concentrations of 0.5, 0.8, 2.0, and 5.0 mM are shown as solid red, green, cyan, and blue lines, respectively. Dashed lines represent extrapolations for each concentration obtained by applying concentration dependent time rescaling factors[1] to the 5.0 mM curve.

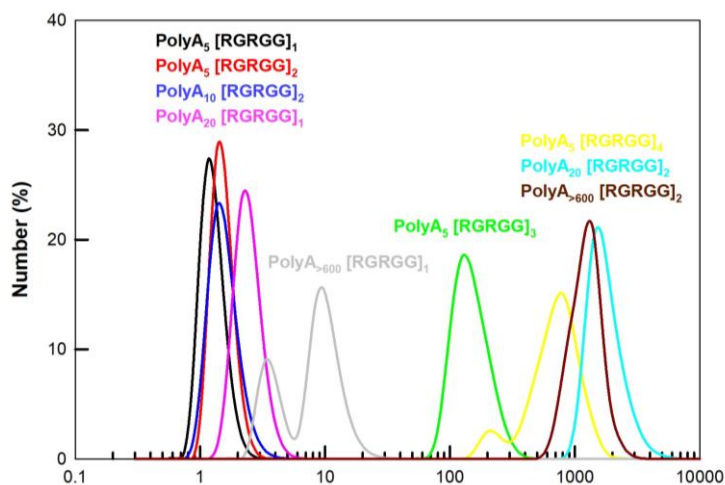
10, 16, 40, or 100 quencher proteins in a cubic box with a width of 321.435 Å, respectively. Remember that while experiments were carried out at a quencher concentration of 840 μM due to solubility limits, atomistic simulations were performed at 5 mM. This time scale correction has been applied in **Figure 6.4.A**.

Figure A5



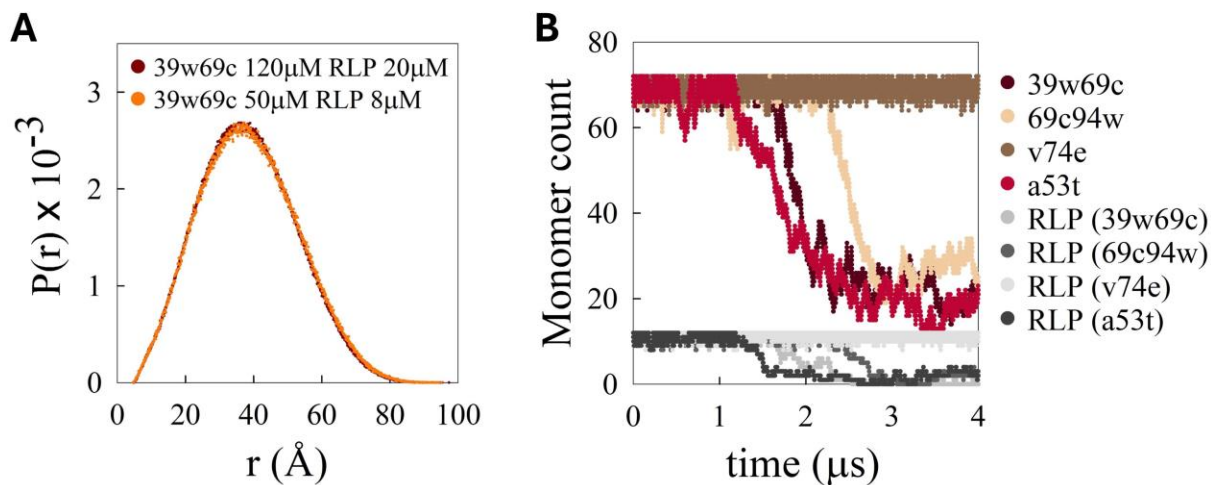
The effect of system size on PolyA_N-[RGRGG]_M LLPS simulations. Simulations were performed using COCOMO 1.2 σ and the same conditions as in **Figure 3.2**. The panels show snapshots after 20 μ s simulations using 100 nm (left) and 200 nm (right) size box systems. We notice that LLPS of our system was not affected by the system size. Writings on top of the panels represent the number of protein/RNA molecules. The depicted systems are PolyA₅ – [RGRGG]₁ (**A**), PolyA₂₀ – [RGRGG]₈ (**B**), PolyA₁₀ – [RGRGG]₂ (**C**), PolyA₁₀ – [RGRGG]₃, (**D**), PolyA₂₀ – [RGRGG]₁ (**E**), and PolyA₂₀ – [RGRGG]₂ (**F**). RNA is indicated in red and protein in blue. A ruler with 10 nm divisions has been added to provide a size reference.

Figure A6



Particle size distributions of peptide-RNA mixtures measured by DLS. DLS counts the number of particles of each size and is indicated here as a percentage of the total counts. Here the size is in terms of the hydrodynamic radius.

Figure A7



Starting conditions and equilibration of the CG systems. **A)** Comparing $P(r)$ of α -syn at different starting conditions of CG simulations. The distance distributions show no apparent difference at high concentrations (red, monomer start, mean = 38.34 Å) of α -syn and RLP compared to experimentally used lower concentrations (orange, pre-formed condensate start, mean = 38.36 Å). **B)** Number of α -syn (red shades) and RLP (grey shades) chains in the dilute phase as a function of run time for monomer started systems. Condensates start to form $\sim 1.5 \mu s$.

Table A1: ITC concentration table. Concentrations of the peptide and RNA in the sample cell after each injection. An injection is 10 μ L.

Total Injection volume (μ L)	PolyA ₁₀ [RGRGG] ₄				PolyA ₂₀ [RGRGG] _x						
	PolyA ₁₀		[RGRGG] ₄		PolyA ₂₀		[RGRGG] _x				
	μ M	mg/ml	μ M	mg/ml	μ M	mg/ml	μ M	X = 1 mg/ml	X = 4 mg/ml	X = 8 mg/ml	X = 10 mg/ml
10	30.80	0.10	1.64	0.00	16.94	0.11	0.91	0.00	0.00	0.00	0.00
20	30.59	0.10	3.27	0.01	16.88	0.11	1.81	0.00	0.00	0.01	0.01
30	30.40	0.10	4.87	0.01	16.82	0.11	2.70	0.00	0.01	0.01	0.01
40	30.20	0.10	6.45	0.01	16.76	0.10	3.58	0.00	0.01	0.01	0.02
50	30.01	0.10	8.01	0.02	16.70	0.10	4.46	0.00	0.01	0.02	0.02
60	29.82	0.10	9.55	0.02	16.64	0.10	5.34	0.00	0.01	0.02	0.03
70	29.63	0.10	11.08	0.02	16.58	0.10	6.21	0.00	0.01	0.02	0.03
80	29.44	0.09	12.58	0.02	16.52	0.10	7.07	0.00	0.01	0.03	0.03
90	29.26	0.09	14.06	0.03	16.46	0.10	7.92	0.00	0.02	0.03	0.04
100	29.07	0.09	15.53	0.03	16.40	0.10	8.77	0.00	0.02	0.03	0.04
110	28.90	0.09	16.98	0.03	16.35	0.10	9.62	0.00	0.02	0.04	0.05
120	28.72	0.09	18.40	0.04	16.29	0.10	10.45	0.01	0.02	0.04	0.05
130	28.54	0.09	19.82	0.04	16.23	0.10	11.28	0.01	0.02	0.04	0.05
140	28.37	0.09	21.21	0.04	16.18	0.10	12.11	0.01	0.02	0.05	0.06
150	28.20	0.09	22.59	0.04	16.12	0.10	12.93	0.01	0.03	0.05	0.06
160	28.03	0.09	23.95	0.05	16.07	0.10	13.75	0.01	0.03	0.05	0.07
170	27.86	0.09	25.30	0.05	16.01	0.10	14.55	0.01	0.03	0.06	0.07
180	27.70	0.09	26.63	0.05	15.96	0.10	15.36	0.01	0.03	0.06	0.07
190	27.54	0.09	27.94	0.05	15.90	0.10	16.16	0.01	0.03	0.06	0.08
200	27.37	0.09	29.24	0.06	15.85	0.10	16.95	0.01	0.03	0.07	0.08
210	27.22	0.09	30.52	0.06	15.79	0.10	17.74	0.01	0.03	0.07	0.09
220	27.06	0.09	31.79	0.06	15.74	0.10	18.52	0.01	0.04	0.07	0.09
230	26.90	0.09	33.05	0.06	15.69	0.10	19.30	0.01	0.04	0.08	0.09
240	26.75	0.09	34.29	0.07	15.64	0.10	20.07	0.01	0.04	0.08	0.10
250	26.60	0.09	35.51	0.07	15.58	0.10	20.83	0.01	0.04	0.08	0.10
260	26.45	0.09	36.72	0.07	15.53	0.10	21.59	0.01	0.04	0.08	0.10
270	26.30	0.08	37.92	0.07	15.48	0.10	22.35	0.01	0.04	0.09	0.11
280	26.15	0.08	39.11	0.08	15.43	0.10	23.10	0.01	0.04	0.09	0.11
290	26.01	0.08	40.28	0.08	15.38	0.10	23.85	0.01	0.05	0.09	0.12
300	25.86	0.08	41.44	0.08	15.33	0.10	24.59	0.01	0.05	0.10	0.12

Table A2: LLPS at different concentrations of peptide and RNA. Imaging from bright-field and confocal microscopy are used in the grid. Concentrations in row two represent the concentration of both the peptide and RNA for each box. The grid is color coded as follows; No condensates (light gray), condensates (dark gray), no condensates 'predicted' (white), condensates 'predicted' (super dark grey), not imaged (black).

	PolyA ₅	PolyA ₁₀					PolyA ₂₀				
(mg/ml)	1	1	0.75	0.5	0.25	0.1	1	0.75	0.5	0.25	0.1
[RGRGG] ₁											
[RGRGG] ₂											
[RGRGG] ₃											
[RGRGG] ₄											
[RGRGG] ₆											
[RGRGG] ₈											
[RGRGG] ₁₀											

Table A3: Residue-specific parameters used in COCOMO, COCOMO2 and COCOMO1.2 σ simulations. Here $\sigma_i = 2r_i \times 2^{-1/6}$, where r_i is the radius of a sphere with equivalent volume of the residue i .

Residue	Mass (amu)	COCOMO and COCOMO2 σ_i (nm)	COCOMO1.2 σ σ_i (nm)	charge	A_i
Ala	71.08	0.253	0.304	0	0
Arg	157.20	0.318	0.381	1	0.87
Asn	114.10	0.281	0.337	0	0
Asp	114.08	0.277	0.333	-1	-0.87
Cys	103.14	0.269	0.323	0	0
Gln	128.13	0.295	0.354	0	0
Glu	128.11	0.292	0.351	-1	-0.87
Gly	57.05	0.233	0.280	0	0
His	137.14	0.297	0.357	0	0
Ile	113.16	0.299	0.359	0	0
Leu	113.16	0.300	0.360	0	0
Lys	129.18	0.306	0.368	1	0.87
Met	131.19	0.301	0.361	0	0
Phe	147.18	0.317	0.380	0	0
Pro	98.13	0.284	0.341	0	0
Ser	87.08	0.261	0.313	0	0
Thr	101.11	0.277	0.332	0	0
Trp	186.21	0.334	0.401	0	0
Tyr	71.08	0.322	0.386	0	0
Val	157.20	0.286	0.343	0	0
Ade	315.70	0.376	0.451	-1	-0.87
Cyt	305.20	0.366	0.439	-1	-0.87
Gua	345.20	0.379	0.455	-1	-0.87
Ura	305.16	0.364	0.437	-1	-0.87

Table A4: Potential parameters (Eq. 6.1) used in MC simulations to match MD simulation probability distributions. The subscripts indicate, ₁parameters applied for distances smaller than the contact minimum, ₂parameters used in the second Gaussian function, ₃the distance at which Debye-Hückel potential becomes effective.

	Villin WT	Villin V10W	Villin K33W	Villin R15T+K30E	SH3
σ [Å]	3.28	3.25	3.42	3.40	3.85
ε [kcal/mol]	6.30	4.80	1.30	2.95	5.00
ε_1 [kcal/mol]	7.00	7.00	7.00	7.00	1.00
potential	12-6	12-6	12-6	12-6	10-5
a [kcal/mol]	0.75	0.45	1.20	1.20	1.20
μ [Å]	4.70	4.52	6.10	5.90	8.00
w [Å]	0.3	0.3	1.9	1.6	2.1
a_2 [kcal/mol]	0.3	0.9	0.5	0.5	0
μ_2 [Å]	6.00	6.85	8.20	4.40	-
w_2 [Å]	0.35	1.40	0.40	0.30	-
d [kcal/mol]	25	11	17	20	0.4
k [Å]	3.5	4.0	4.0	4.0	60
Cutoff ₃ [Å]	10	9.5	10	9.5	10