5-2024

Intrinsically Disordered Proteins and Their Role in Biomolecular Condensates

Danielle Latham
Clemson University, drlatha@clemson.edu

INTRINSICALLY DISORDERED PROTEINS AND THEIR ROLE IN BIOMOLECULAR CONDENSATES

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Physics

---

by
Danielle Rebecca Latham
May 2024

Accepted by:
Dr. Hugo Sanabria, Committee Chair
Dr. Sapna Sarupria
Dr. Feng Ding
Dr. Kasra Sardashti

ABSTRACT

Proteins are biomacromolecules responsible for the functions of life. While classically proteins are thought to be well structured in order to perform a specific function, 50% of proteins within Eukaryotic cells contain intrinsically disordered regions (IDRs), regions with no well-defined structure. IDRs are often used for cell signaling, responding to external factors such as temperature changes or the presence of small molecules. To understand how IDRs can function without structure, it is important to understand the dynamics of such systems. Understanding IDR intramolecular and intermolecular interactions will shed light on IDR dynamics. Intramolecular interactions are first explored using fluorescence spectroscopy methods and polymer modeling using the C-terminal domain (CTD) of GluN2B as a sample IDR. Intermolecular interactions are explored using coarse grain simulations to predict biomolecular condensation formation for protein sequences with varying numbers of glutamine. It is found that IDRs show complex dynamics in which secondary structure may play a larger role than previously expected.

DEDICATION

To Sofie, whose codependency knows no bounds. You have been there every step of the way. In moments of doubt and exhaustion, your not-so-gentle presence has provided a comforting reminder of the importance of food (mostly yours) and taking breaks to enjoy staring out into the backyard. Your intuitive nature and keen awareness of my emotional state have offered solace during times of stress, often unnoticed by others. Your loud calls have offered reassurance when desperately needed and distractions often in the worse moments.

Our unwavering bond has instilled in me a sense of responsibility and purpose, driving me forward with the knowledge that your well-being and happiness are intertwined with my own. Your demanding antics and forceful gestures have brought moments of levity to the most challenging of days, serving as a reminder that you are the sole important thing in my life, according to you.

As I embark on the next chapter of my academic journey, I am deeply grateful for your unwavering presence and unwavering support. Your steadfast companionship has been a source of strength and inspiration, enriching my life in ways beyond measure. I cannot imagine my life without you. For this reason, I will drag you halfway around the world. We were truly meant to be together.

ACKNOWLEDGMENTS

I would like to express my deepest thanks to Dr. Hugo Sanabria. He has gone above and beyond to ensure I achieve my best and aim to live up to my full potential. Further appreciation goes to my committee members: Dr. Sapna Sarupria, Dr. Feng Ding, and Dr. Kasra Sardashti. Thank you all for guiding my dissertation and taking the time to read this very long document. Special thanks to Dr. Sapna Sarupria who lead my learning and discovery of LAMMPS. I never expected to have computational research, but the world had other plans. Your positivity has greatly helped on this adventure.

Many people have assisted me in my journey to get this far. Thank you to Inna Yanez-Orozco, George Hamilton III, and Narendar Kolimi for having the patience to introduce me and train me for all the basics required to work within the wet lab. Particularly Inna, who was always there when I had a question or to discuss a curiosity. Thank you to Exequiel Medina for training me on setting up single molecule experiments on the confocal. You would always answer my messages despite the time or how often I asked the same question.

Brian Sanders and Frank Duffy, thank you for taking the time and energy to help me, an experimentalist at my core, understand python coding enough to process my data and create all the graphs required. It is because of your guidance that I can now code rather well in python on my own.

The projects listed in this dissertation would not have been possible without, Dr. Mark Bowen who supplied the N2B plasmids, Nabanita Saikia who ran the DMD

simulations for N2B, Krushi Patel and Sterling Nicholas who helped purify and label the N2B samples, and Braden Holst who assisted with initial simulations in LAMMPS.

A big thank you to all the undergraduate students I have interacted with at my time at Clemson University. From the students within the physics labs to the REU students, interaction with you have shown me my passion for teaching. This has led to the personal goal to expanding biophysics education and exposure.

On a more personal note, I would like to thank my many lab mates, particularly Inna Yanez-Orozco, Anay Fernanda Lazaro and Brian Sanders who have always been around for good conversations about science, life, and anything else of interest. You three have reminded me to take time to look up and enjoy the world around me. It has been such a wonderful experience working with Krushi Patel and Sterling Nicholas. You guys have really kept me on my toes and constantly remind me the importance of understanding good basics. Thank you for all your wonderful questions and ideas.

TABLE OF CONTENTS

Page

Table of Contents (Continued) Page

LIST OF TABLES

List of Tables (Continued)

Table Page

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE

BIOLOGICAL SYSTEMS

Proteins: Structure and Function

DNA is the blueprint of life. It is through the central dogma that DNA becomes RNA which in turn becomes proteins. At the core, DNA is made up of four different types of base pair that code for proteins. RNA is also made up of four different types of base pairs that come from a one-to-one transition from DNA. However, proteins are biomacromolecules that function as nano-machines to keep life living. Protein have complex tasks that require more variety than DNA or RNA. With increased complexity, proteins are made up of 22 different amino acids which vary by the attached functional group. Amino acids production occurs through a RNA polymerase, which reads RNA codons to produce the associated proteins. Codons are comprised of three RNA base pairs and code for a single amino acid. All amino acids share an Amine-Carboxylic acid backbone which allows them to linearly chain through covalent bond through the nitrogen on the amine group and carbon on the carboxylic group leaving a water molecule as a biproduct through a dehydration synthesis reaction. Proteins widely vary in length with insulin (a small peptide) having only 54 amino acids to proteins in the Titin family which have over 30,000 amino acids (1,2). This wide variety in length allows the 22 amino acids to form very complex and different proteins the same way the 26 English letters can spell countably infinite words. Proteins are therefore read along the backbone

from the amine side (N-terminal) to the carboxylic side (C-terminal) to give the order of amino acids, commonly referred to as the protein sequence.

Classically, proteins are thought to have a defined structure which relates to their function, such as enzymes which are typically viewed as a lock and key models where interactions with small substrates into a binding cavity lead to a structure change of the enzyme (Figure 1.1). However, protein structure is more complex than a simple structure and occurs within a structural hierarchy. The protein sequence gives the primary structure of the protein. As functional groups of nearby amino acids interact with each other due to proximity, hydrogen bonding and electrostatic interactions lead to commonly occurring secondary structures including alpha helixes and beta sheets. Alpha helixes are a helix structure stabilized through hydrogen binding, while beta sheets are sheet-like structures also stabilized through hydrogen binding. Together, the localized secondary structures come together to form the full three-dimensional, tertiary structure of the protein. Often a protein will interact with other proteins of the same (homo) or different (hetero) type to form larger, multi-protein complexes. Such complexes are referred to as the super-tertiary or quaternary structure. Understanding the relationship between protein structure and the corresponding function of the protein is of upmost interest. Active research is also working to determine predictive mechanisms for determining a protein's tertiary structure given its primary structure.

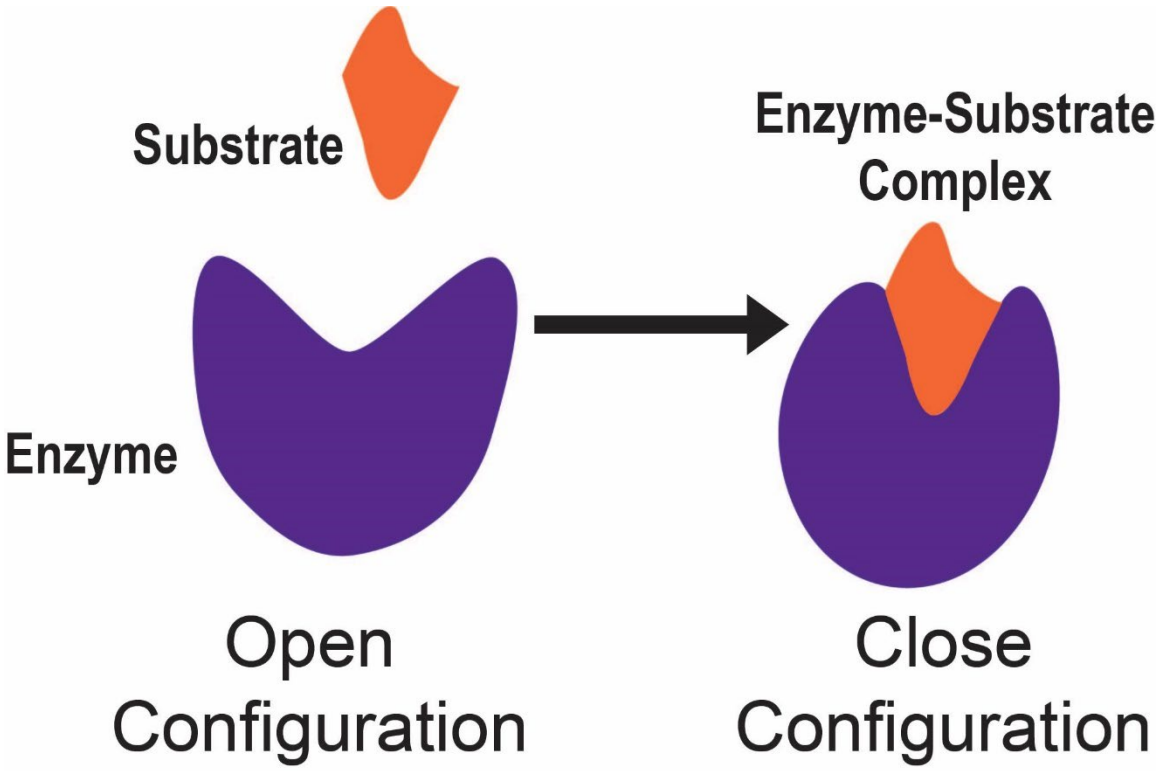


Figure 1.1 **Lock and key model for enzymes.** The enzyme (purple) remains within an open configuration allowing the substrate (orange) to access the binding location. Once bound, the enzyme changes configuration to a closed state.

Many proteins contain several tertiary structures or will change tertiary structure depending upon the environment. In this case, the tertiary structure is referred to as a conformational state of the protein. The easiest thought exercise is a simple enzyme which starts in an 'open' state. Once a substrate binds to the enzyme, it will change conformational states resulting in a 'closed' state. In this situation, the enzyme protein has two different conformational states and thus, two tertiary structures (Figure 1.1). However, many proteins have several conformational states depending upon the local environment or just stochastic movement due to thermal fluctuations. Each conformational state contains its own enthalpy level and is separated from other configuration states by an energy barrier. Energy landscapes can be created for proteins where stable configurations exist as basins with saddle points connecting exchangeable configurations (3-6). The classical view of protein sequence to structure to function paradigm lacks a necessary dynamic aspect to proteins (7-10).

Intrinsically Disordered Proteins

The importance of protein dynamics was not a large research focus until the late 1990s with the discovery that not all proteins require a three-dimensional structure to function (11,12). Some proteins are intrinsically disordered; proteins which rapidly sample configuration space to function. These intrinsically disordered proteins (IDPs) are highly dynamic with no set secondary structure. Many proteins will contain structured regions, also referred to as domains, separated by intrinsically disordered regions (IDR).

Nowadays, it is thought that over 50% of all Eukaryotic proteins contain an intrinsically disordered region (13,14), with IDRs varying in length depending upon the protein.

The dynamic nature of IDPs makes disordered proteins difficult to study with historical methods such as X-ray crystallography or Cryo-Electron Microscopy depending upon a protein having a well-defined structure to obtain good measurements for analysis. Since IDPs are hard to study, current research is biased to well-structured proteins (15). With fast dynamics occurring within the nanosecond range, it has been difficult to obtain good measurement regarding the dynamic nature of IDPs. Recently advancing technology has allowed for faster time resolution by introducing fluorescent based experiments for protein study. Such methods include Fluorescence Correlation Spectroscopy (FCS) and Förster Resonance Energy Transfer (FRET), which are discussed in more detail within Chapter Four.

IDPs and IDRs have been shown to have important roles within the cell, particularly involved in signaling (16) and DNA interactions (13). The disordered nature makes these protein regions more responsive to environmental changes making them perfect for signaling and regulation (14). Even though there are many open questions regarding IDRs, it is known that they are required and necessary for several proteins to function correctly. Recent advancements of detection methods will give insight into the importance of IDPs including their role in biomolecular condensate formation.

Biomolecular Condensates

It is common for a certain research topic to be explored several different times as technology evolves allowing for a deeper investigation and understanding. In neurons, dark granules were first noticed in the 1950s when electron microscopes showed very dark, electron dense dots along the membrane (17). This region came to be referred to as the Post-Synaptic Density (PSD) due to the large density of electrons location at the Post-Synaptic Terminal. At the time, little technology existed to gather a better look. In the 1980s, it was found that certain combinations of proteins and molecules when mixed in vitro would appear cloudy or cause droplets to form. But the technology required for a deeper look was still lacking. With the advancement of fluorescent measurements and detection devices, these mysterious results could be further explained.

In 2009, Brangwynne published a research article which used fluorescence recovery after photobleaching (FRAP) showing the dark granules within the electron microscope contained liquid-like properties including dynamics and fusion (18). Brangwynne postulated the dark spots form through a condensation-like mechanism resulting from a demixing phase transitions within fluids. As more research was completed and similar systems were discovered, the term membraneless organelles were termed to describe such systems, since the result is a stable, liquid-like droplet separated from the cell cytoplasm without a surrounding membrane. The formation and stability of the droplet are thought to form via liquid-liquid phase separation (LLPS) (19, Wei, 2017 #82,20).

With this new viewpoint of LLPS, older experiments were reevaluated with past fluorescence experiments such as Fluorescence Recovery After Photobleaching (FRAP), Time-Lapse Fluorescent Microscopy (TLFM), and Fluorescence Loss in Photobleaching (FLIP) showing the cell nucleus has sub-compartments with liquid-like mobility (19,21). It is shown that the liquid-properties within the nucleoli define the characteristics including shape and size of the nucleus (22). This evidence suggests the cell has different methods for organization, adding membraneless organelles to the many already known lipid surrounded organelles such as the nucleolus and mitochondria. Currently, the term membraneless organelles has been phased out in preference to the term biomolecular condensates, referring to the material (bio), size (molecular) and manner of formation (condensation) (23). Biomolecular condensates are highly dependent upon environmental factors such as salt concentration, temperature, and pH, similarly to LLPS (24-26).

Biomolecular condensates are found in many locations within the cell, performing many different functions and roles. Liquid-liquid phase separation as a driving force for biomolecular condensates explains many different aspects of life including efficient organizing in the nucleus (27), mRNA storage with ease of access (28), sorting between double stranded and single stranded DNA (29), capturing misfolded proteins and aiding in their degradation (30-32), and heat stress resistance in embryos (33). While early experiments focused around identifying biological systems that may be classified as biomolecular condensates, the focus has shift to understanding the behavior of molecules which construct as well as form biomolecular condensates. Biomolecular condensates show local, short-range order with long-range disorder (34), with evidence pointing to

weak electrostatic interactions as a driving force for phase separation (35,36). Weak interactions are required to keep proteins within a biomolecular condensate highly dynamic while exchanging biomolecules with the cellular cytoplasm (35,37-39). While biomolecular condensates seem essential for life, there is still much that is unknown.

Many proteins that have been found to drive LLPS are intrinsically disordered proteins (IDPs) or contain intrinsically disordered regions (IDRs) (40-42). Since IDPs are less likely to form crystal structures, they do not have as strong of a liquid-solid competition (43). For proteins with structured domain attached by an IDR linker, the linker's property can determine if phase separation occurs (44). Further, proteins can phase separate by IDRs while leaving the structured domains intact (45). Biomolecular condensates places proteins in locally high concentrations increasing the interaction probabilities between domains, enhancing the formation of dimers or higher order complexes (46). While some IDPs phase separate on their own, others require the presence of multiple proteins or RNA (47,48). Biomolecular condensates can even have a varying composition along radial axis resulting in the formation of biomolecular condensates within biomolecular condensates (49,50). As not all IDRs can phase separate, it becomes important to develop predictive methods for biomolecular condensation.

The proteins that phase separate contain several interactions sites per protein (16,51,52). The number of interactions cites available per protein is referred to as the multivalency of that protein. To understand the importance of multivalency, one must understand the interactions between amino acids. Due to the varying functional groups

which construct amino acids, certain amino acids are attracted to other amino acids, while some amino acid combinations have little to no interaction. The attractive amino acids can be thought of as stickers while the amino acids who interact very little can be thought of as spacer. It is the multivalence property of IDPs allows them to act as scaffolding for biomolecular condensates and leads to the sticker and spacer model (49,53,54). However, the sticker and spacer model is not completely accurate and does not explain all of the physics behind the formation of biomolecular condensates (47).

Advancing the sticker and spacer model by allowing each amino acid to have interaction parameters with all other amino acids giving a range of interaction values rather than being simply assigning a sticker or spacer value. This allows different functional group interaction within amino acids to be considered. Evidence shows Pi-Pi interactions between amino acids with aromatic rings being the driving force and most important interaction for phase separation (55,56). Computational force fields are being developed giving more weight to Pi-Pi interactions while trying to determine the driving forces within biomolecular condensate formation.

While biomolecular condensates are stable with liquid properties, their liquid properties can change over long-time scales resulting in maturation into amyloids or hydrogels leading to protein aggregation resulting in disease (57,58). In vivo experiments show biomolecular condensates ‘age’ and change over time, becoming more ‘gel-like’ and stiffer with droplet shrinks over time frames as short at eight hours leading to the fluid properties acting like Maxwell fluids instead of classical Newtonian fluids (59). Though the cell contains safeguards to help assist in breaking down misbehaving

biomolecular condensates (60). Better understanding biomolecular formation and properties will aid in understanding how they lead to aggregation diseases. Furthermore, a deeper understanding may result in engineering biomolecular condensates in a manner that can be manipulated to capture toxic proteins or aid in healing.

Predictor Of Naturally Disordered Regions (PONDR)

It becomes necessary to have a quick method to determine intrinsically disordered regions (IDRs) within proteins. Many still believe in the sequence to structure paradigm, thus Predictors Of Natural Disordered Regions (PONDRs) are designed to determine protein disorder based solely on sequence (61, Dunker, 2002 #371). IDRs are formed largely from low complexity domains (15) and Shannon entropic factors (62). It was found that Neural Networks are slightly more accurate at predicting disordered regions than logistic regression or discriminant analysis (63). Therefore, there are currently several Neural Network modeling methods used to predict disordered protein regions (64). Neural Networks are trained for predictions using known data sets. Due to the historic nature of protein structure research, special care must be taken to ensure there is no bias within the training set used for neural networks. However, the accuracy of every method depends upon the type of disordered protein that is trying to be predicted (65).

There is still debate regarding what is defined as disordered (66) with the most common answer being proteins that lack a set 3-D structure. Each protein has differing levels of disorder or disordered regions which may change over time or respond to various stimuli. Some proteins have disordered regions that form a structure upon binding

(67). But many proteins have two or three stable configurations and are not considered disordered. While it is easy to define the start and end of an alpha helix, it is difficult to accurately predict a clear boundary between ordered and disordered regions (68). There are big differences in the experimental definition for disordered regions which make predictors who need these data set less accurate (69). However, PONDR is still being updated and improved (70) with current models being up to 84% accurate (14,70). Since PONDR VSL2 combines a long (>30 residues) and short (<30 residues) length predictor to obtain all possible disordered regions it will give the best predictions for our samples of interest (71). However, PONDR VL-2V method is known to be the strongest predictor with highest accuracy in identifying longer disordered regions (65). There continues to be advancement and development in IDR predictors, such as Rapid Intrinsic Disorder Analysis Online (RIDAO) (72) leading to advances and increase interest in the field of disordered proteins.

Protein Polymer Models

Since proteins are made up of amino acids, often referred to as residues, polymer models can be applied to proteins, in which the proteins are thought of as polymers constructed from amino acids, which serve as monomers. Different polymer models are used for different types of proteins with Gaussian chain models working well for structured proteins and worm-like chain model working well for IDPs (73). Protein folding can be decently explained by collapsed polymer modeling when a heteropolymer is used where each amino acid is represented by a different monomer with its own

properties (74,75). This collapsed polymer modeling does a good job at predicting the radius of gyration for well structure proteins (76). However, as intrinsically disordered proteins (IDPs) do not fold into a set 3-D structure, collapse polymer models are not enough to model all proteins leading to guided polymer models based on statistical mechanics considerations and free energy landscapes (77).

Worm-like chain (WLC) models are often used to explain the rapid dynamics of IDPs (78). WLC models use the end-to-end distance distribution of polymers to classify the stiffness of the polymer through a persistence length. The persistence length is a measure of distance along the polymer before directional correlation is loss. The larger the persistence length, the stiffer the polymer. Since WLC models are only dependent upon a distance distribution, experiments such as Förster Resonance Energy Transfer (FRET), which obtain a distance distribution between labeling sites, offers a perfect method to probe the system (79). Often end-to-end distances are obtained between several locations along the protein, allowing a network of labeling locations with various distance distributions to be determined. Often labeling locations contain overlapping regions, referred to as segments (Figure 1. 2). Each segment will have an average stiffness related to the persistence length. Modeling the system as springs in series, the values obtained from the probed systems can be analyzed to determine the stiffness of each segment.

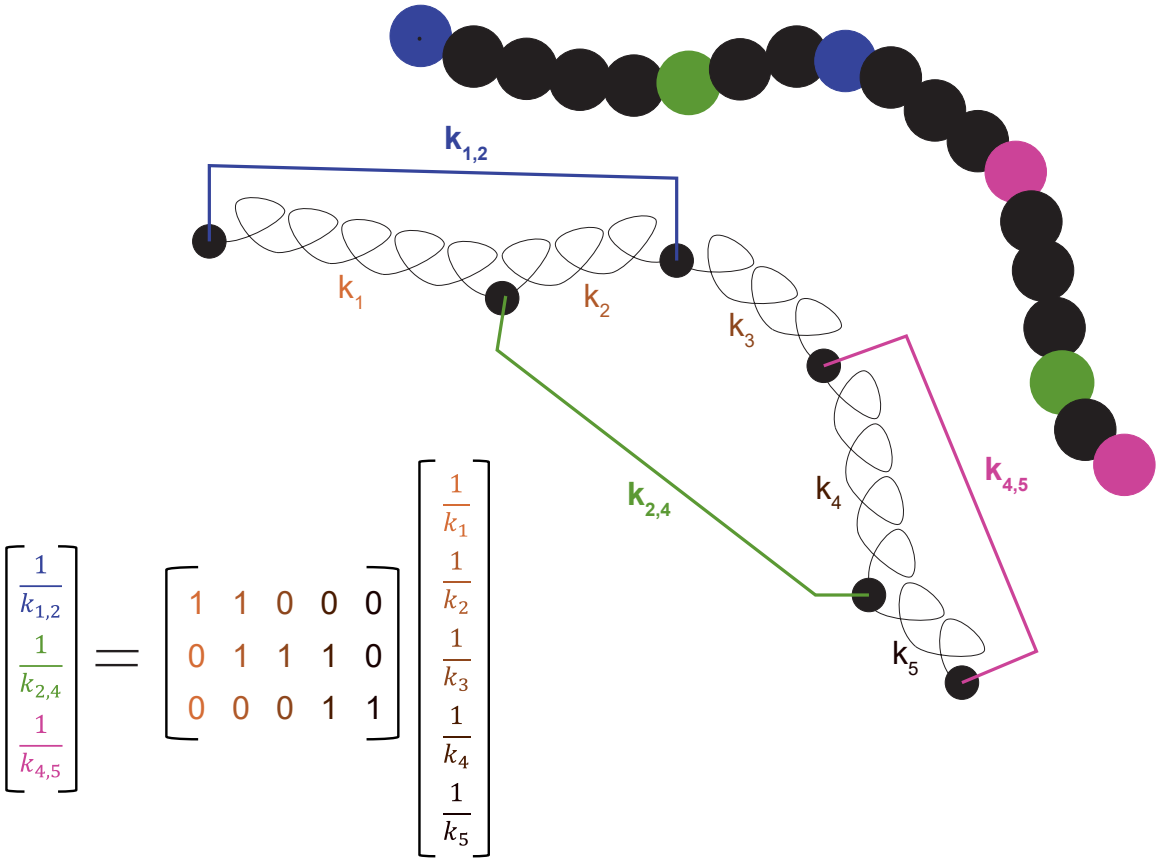


Figure 1. 2 **Example of an elastic network model for a polymer.** Polymer is shown in black with labeling locations shown in blue, green, and pink. Stiffness, k, for each labeling location matches the color of the labels. Segment numbers are shown in varying shades of orange. Spring in series equation leads to the interaction matrix shown in the bottom left corner Color coding of matrix equation matches that of the figure.

There are many considerations that can be added to polymer models, such as the thickness of the polymer, which improves models for certain proteins (80). Models can also be applied to experimental measurements to help understand the physics within the system. Rouse model adds internal friction to the system which can expand knowledge on amino acid interactions, particularly regarding IDPs (81). Polymer models have been used to explain why water is a bad solvent for proteins and why proteins require salts and buffers for stability (82). Even lattice-based polymer models offer a powerful tool for studying protein dynamics and structure prediction (83). By combining polymer models with clustering from molecular dynamic simulations, several dynamic protein phenomena such as meta-stable states, transition states, and stochastic dynamics can be explained (84). Some polymer modeling can predict posttranslational modifications such as phosphorylation based on electrostatic distribution and stability changes caused by the modification (85).

Proteins work so well as polymers that they are being used to replace more complex, harder to control polymers. Since proteins can be harvested through biological expression systems, they become easier to control through their DNA gene code (86). Proteins are starting to replace typical polymer in drug delivery (87) and used more often in commercial locations such as fabrics and material (88).

Prions

Historically, illness and diseases were thought to only spread by bacteria or viruses. The idea that an incorrect protein configuration could be the difference between

healthy and diseases cells was a controversial and novel idea at the time. The 1997 Nobel Prize in Physiology or Medicine was awarded to Stanley Prusiner for his work showing that scrapie in sheep was caused by infection of a misfolded proteins (89). He coined the term prion to refer to such infectious proteins, describing prions as small proteinaceous infectious particles which are resistant to inactivation by most procedures that modify nucleic acids (90). Prion diseases come from a change in protein's tertiary structure, its amino acid sequence remains unchanged (91).

The term prion relates to a family of proteins which lead to various transmissional diseases. Prion disease includes scrapie in sheep, Bovine Spongiform Encephalopathy (Mad cow disease), Chronic Wasting Disease in Elk/Deer, and Creutzfeldt-Jakob disease in humans (92). While prion diseases are found in various species, there is a species barrier for transmission (93); for example, Mad Cow disease cannot transmit from a cow to a sheep, while scrapie does not transmit from sheep into cows. Prion diseases behave differently between species (94). However, all prion diseases develop with a neurological component and aggregation within brain tissue (95,96). The disease-specific configuration often has the same post-translational modifications as the healthy prion (97), giving further evidence for the importance of protein configuration in relation to protein sequence.

Many essential proteins within the cell contain regions with properties similar to prions. These prion-like domains are low-complexity, disordered regions with a high concentration of glutamine and asparagine (Q/N-rich), which lack hydrophobic residues (98). The composition of the prion-like domain further affects aggregation formation with

structure and dynamics being dependent upon number of glutamines (Q) in the sequence. Sequences with a higher glutamine percentage have more loops and bends leading to higher-order aggregate morphologies (35,99). Furthermore, as the length of polyQ regions increase, aggregation rates increase becoming larger and irreversible (100).

Aggregation formation is dependent upon secondary structure. Glutamine can act as an ambivalent hydrophobe to help the formation of alpha helixes (101). Typically, these alpha helixes form coiled-coil regions which can lead to beta-sheet amyloids (45). These beta-sheets can be toxic if a large enough concentration is built up, but coiled-coil interactions may assist in stabilizing the aggregation away from toxic situations (102). There are many ways to form beta-sheets, evidence shows that Q-rich regions may act as polar zippers that seed aggregation (103). There is evidence showing polar zippers lead to parallel beta sheets (104). With more evidence for parallel beta sheets arising during aggregation on Q-rich regions with no change on the effect occurring when Q-rich regions are shuffled (105).

Not all coiled-coil interactions lead to beta-sheet aggregation. The family of type III Secretion systems injectosomes uses coiled coil interactions between intrinsically disordered protein regions to help form networks required for function (106). Also, increasing the coiled coil interactions within a Q-rich region results in a decrease of protein concentration within a biomolecular condensate. Decreasing coiled coli interactions leads to an increase of concentration leads to the belief that disorder is the most important factor in biomolecular condensate formation (107). However, more recent simulation work shows that the beta sheet formation and interactions may be more

important for biomolecular condensate formation than previously thought coil-coil interactions. (108).

While prions are most widely known for their disease-causing capability and toxicity, prions have beneficial properties as well (109). There is a large amount of evidence that shows many prion-like domains promote biomolecular condensate formation (110). Prion-like domains can act as stress responses that sense physio-chemical changes in the cells and form biomolecular condensates as protection (111). Evidence shows prion-like domains act as scaffolding which hold biomolecular condensates together (98,112,113). The majority of proteins with prion-like domains are transcription factors, proteins that bind with DNA to control the rate of transcription occurring within the cellular nucleus (98,112). In RNA binding proteins with a prion-like domain, the prion-like domain helps promote phase separation while RNA-binding domains control the dynamics and properties of the phase separation (114). Without the RNA-binding domain, phase separation still occurs with the prion-like domain being sufficient to drive it. Showing the physiological relevance of prion-like proteins, research shows some polyQ regions engaging in phase separation assist with cargo transport (115), while plants use prion-like domains for temperature dependent phase separation to signal blooming (116). Saccharomyces cerevisiae stress recovery due to starvation and acidic conditions is dependent upon prion-like, Q-rich, low-complexity domains which help pump protons across the cell barrier (117).

However, it is difficult to predict properties of prion-like regions; for example, Transthyretin (TTR) aggregation is enhanced by the presence of Q-rich protein Sup35

when in the diseases configuration but is not affected by the healthy configuration (118). The same study revealed that PIN+, another Q-rich prion-like protein, does not affect TTR aggregation. Many aspects of prion structure and interactions differ from expected behavior. For example, when forming beta sheet aggregates, prion peptides place tyrosine outwards; against common thoughts (119). The surrounding regions and domains play an important role in the physical properties of prion-like domains and their aggregation formation, size, and location. Some prion-like domains contain side domains that are proline(P)-rich and reduce the toxicity of polyQ regions within the cell (120,121). Introducing small Q-rich peptides to known toxic aggregative prion-like domains make the aggregation less toxic to the cell (122).

While prion-like domains are known to become toxic, there are safeguards in place to help prevent and deal with aggregation before diseased prions can cause damage to cells. Mutation studies with prion-like domains show decreases in hydrophobicity leads to liquid like droplets near the nucleus with higher toxicity while increasing hydrophobicity leads to aggregation away from the nucleus with lower toxicity (123). The decrease in toxicity may be due to trapping of toxic protein by aggregation away from the cellular nucleus. There is evidence that shows Q-rich prion-like domains can capture, isolate and trap toxic polyQ aggregates within control compartments formed by phase separation (124). Further, the Q-rich droplets suppress the polyQ toxicity by leaving a nearly non-existent polyQ concentration within the cellular cytoplasm, the dilute phase of the phase separation (125). Chaperones proteins have been shown to

correct or alter condensates that contain mutations or have the wrong physical properties (126). This helps prevent toxicity from spreading and avoids diseases.

Despite these safeguards, prion-like diseases still occur. Once an infectious prion is introduced to the system, it interacts with the hosts native prions introducing an intermediate folding state which changes the free energy leading to misfolding (127). In some cases, biomolecular condensates can lead to aggregation of prion-like domains, since the new folded state can lower the energy threshold from liquid to solid phase transitions (128,129). However, crowded cell environments and other experimental limitations have restricted the accessibility of in vivo studies. In vivo experiments show rates of prion multiplication and aggregation orders of magnitude slower than the rates found by in vitro experiments. There is work to be done to create a mapping between in vitro and in vivo experiments regarding the rate of aggregation and multiplication for prion-like proteins (130).

There are many human neurological diseases which present with a similar mechanism to prion diseases, referred to as prion-like diseases. Huntington, Alzheimer's, Parkinson's, and Amyotrophic lateral sclerosis (ALS) are prion-like diseases, caused by insoluble aggregation of typically highly dynamic, disordered proteins within the brain (96). There has recently been a shift in the research of these aggregation mechanisms to couple A- β aggregation with prion diseases referring to the group as prion-like diseases (131). This is due to A-β aggregation occuring through a prion-like mechanism (132,133), where a single misfolded protein will alter those around it leaving to a cluster of misfolded proteins of all the same type which forms aggregates. Further, prion-like

diseases can spread through intercellular transfer (95,134). This leads to a few diseased prions moving between cells, allowing for a diseased cell to infect healthy cells within the same host. Prion-like diseases currently have no cure or treatment. Since cells can be infectious with just a few prion-like proteins, but symptoms don't occur until after prion aggregation starts (135), prion-like diseases aren't typically discovered until after it is too late for medical care for the patient.

Despite the decades of prion research and a century of investigating aggregation, there is still a lack of knowledge and understanding regarding prion-like diseases (136-138). The two fields are closely related and can learn a lot from each other. As in vivo experimental techniques advance, it becomes crucial to understand the connection between in vitro and in vivo experiments (139). The phase separation properties of prion-like domains has just started to be expanded up on, but already show extreme importance to the proper functioning of cells and thus life.

Transcription Factors

Transcription factors are a class of proteins that regulate the first step of gene expression. This first step is known as transcription, when DNA is read to give the associated mRNA, which in turn produces the associated protein. Control of this process is necessary for cellular survival, helping indicate when to advance within the cell cycle. Furthermore, for multicell organisms, all cells contain the same core DNA, but cells manage to specialize in creating different cell types through a process referred to as differentiation (140). Cells differentiate by altering the concentrations and types of

expressed proteins within the cell. Transcription factors can act as activators which promote transcription of a specific gene leading to increase in the associated protein concentration, or as repressors which limit transcription thus decrease protein concentration. This regulation of transcription is necessary for several biological functions including cell differentiation, cell cycle progression, and cellular stress response.

Due to the important role of transcription factors, there is much active research on the topic. There are currently over 1,600 known human transcription factors with large databases constantly being updated to reflect newly gathered information (141,142). A large interest is placed in understanding which transcription factors control which genes. For transcription factors to alter transcription, they bind to DNA through DNA binding domains which target DNA motifs, short DNA sequences often six to twelve base-pairs in length (140). Different transcription factors target different DNA motifs with over 100 known eukaryotic DNA binding domains (140). Due to the large nature of the human genome, identifying all DNA motifs is a difficult task improved by current advancements in artificial intelligence tools (140,143).

Most transcription factors must work in complexes such as dimers or higher ordered multimers to bind to DNA (144). Often, the same transcription factor can be found in more than one complex, varying in structure, binding partners and/or stoichiometry. The differences between complexes affect the behavior of the transcription factor allowing certain transcription factors to behave as an activator or repressor depending upon the presence of coactivators or corepressors (140). Transcription factors

are highly dependent upon the local environment including the presence of other transcription factors, small ligands such as hormones, and temperature (144). This allows transcription to respond to outside stimulus. Such changes can signal cell differentiation or progression within the cell cycle. Transcription factors are also responsible for biological age advancement such as puberty due to transcription factors response to hormones (145) as well as triggering cell death due to toxicity or lack of resources (146,147). The different functions of brain lobes can be explained by neuron differentiation through changes in transcription factors between neurons(148-150).

The majority of DNA is non-accessible to standard transcription factors with only 10% of the human genome being actively available with 12% being open and weakly used (151). The remaining 78% are silent, packaged up within chromatin. However, during different aspects of the cell cycle and during differentiation, different chromatins of DNA are required to be accessed. A class of transcription factors, referred to as pioneer transcription factors, can bind to and interact with DNA packaged along silent chromatin (151,152). Pioneer transcription factors can open chromatin and facilitate the dissociation of histones allowing access to the previously inaccessible DNA allow expression of a different set of proteins (151, Lu, 2023 #501). As more DNA is exposed to allow for a change within the cell, transcription factors can also promote chromatin occupancy of DNA, winding up DNA which is no longer required at the current time (153). This shows transcription factors have a variety of functions and roles within biological systems. Experimentally studying such systems remains difficult due to their complexity , but advancements are being made (154).

Transcription factors are responsible for regulating many aspects of cellular function for events that affect the cell both internally and externally (155). Due to the wide roles performed by transcription factors, it is not surprising that they are linked to several disorders. While the mechanisms are still widely unknown and complex, majority of linked disorders are neurological and developmental such as autism, Parkinson's, schizophrenia, bipolar, and other psychiatric disorders (156-159). Transcription factors have been linked to cancer due to their role in cell apoptosis (160). This makes transcription factors a possible route for cancer treatment. However, upstream and downstream affects make determining the exact role of transcription factors within the disorders difficult.

With the large importance of transcription factors, how are transcription factors regulated? Some transcription factors are regulated by an autoregulatory circuit where the transcription factor regulates its own gene (161). Such systems can be complicated with many different regulatory steps where the protein level could be decreased such as RNA splicing, mRNA degradation or a halt signal sent at translation (process of reading RNA to protein). Further control can come from a ligand dependence of activation or requiring a higher order complex for the transcription factor to bind. Some transcription factors have strong response to external stimuli such as stress, injury, infections, hormones, growth factors all of which can stimulate or decrease the transcription factor function depending upon the factor and the transcription factor in question (145). Other proteins can act as binding partners with transcription factors, which drastically alter the transcription factors behaviors (160). In summary, the process of transcription factor

regulation is very complex with many factors creating a largely open research field (145-147,160).

Transcription factors are highly dynamic with majority of transcription factors containing an intrinsically disordered region (IDR). At least 85% of transcription factors contain some measure of IDR (162), with the complexity of species increasing with the length and variety of IDRs. (163). The structure of most transcription factors results in ordered DNA binding domains surrounded by IDRs (164). While transcription factors properties are dependent upon the disordered regions, there are still debated hypothesis on why (165). Evidence shows ligands change transcription factor behavior by altering the IDR within the transcription factor allowing for different functions or binding affinity (166). Other studies show the IDR may serve as a secondary identifier for DNA binding sites where the IDR rapidly searches for a specific DNA motif through weak interactions allowing the transcription factor to get close to selected motifs (167). This allows the DNA binding domain to quickly localize to the binding motif. In this case, the IDR supplies kinetic proofreading, explaining the specificity with which transcription factors bind some possible binding sites while leaving most other sites unoccupied (168). The IDRs of transcription factors are also shown to offer a method of regulation for the cell. Posttranslational modifications such as phosphorylation can alter the dynamics of IDRs, thus altering the behavior of the transcription factor such as no longer allowing certain binding partners or changing the allowed binding partners. Phosphorylation can be triggered by biochemical signals from cell-surface receptors altering the function of transcription factors (145). Further evidence shows the location of the phosphorylated site

can have different affects upon the transcription factor allowing for more precise regulation (169).

Transcription factors must function in crowded environments within the cell nucleolus and are known to form clusters of highly dense protein (170). These clusters have been linked with similar properties to biomolecular condensates (171). The IDRs within transcription factors are thought to drive phase separation leading to the formation of biomolecular condensates. Evidence shows that the DNA binding domain also plays a role in clustering, possibly in moving the biomolecular condensate to the appropriate location inside the nucleolus (170). Placing engineered transcription factors that form biomolecular condensates upon activation with light into mice has been shown to control transcription. When light is present, gene expression increased up to five times for targeted proteins (172).

CHAPTER TWO

ENERGY AND MIXING

Energy and Entropy

Various energy sources exist all around us, including within our own bodies. Without understanding energy, one cannot fully understand biological systems. Energy is what keeps us alive and moving forward. Biological systems function in a manner that minimizes energy required, leading to the most efficient systems possible (173). Thus, most biological questions can be explained and solved by energy minimization. This includes protein folding and dynamics, DNA folding and storage, cell cycle and propagation, as well as the formation of biomolecular condensates. Thus, thermodynamics is highly important for understanding biological systems. Before diving into the complexities involved in biology, it is important to understand basic thermodynamics starting with energy and entropy.

The environment exchanges energy (U) between systems via heat (Q) and work (W), leading to the first law of thermodynamics: $\Delta U=Q+W$. Heat is an addition of energy through changes in temperature (T) and entropy (S). Therefore, a system where no work is performed, there is a balance of energy and entropy at equilibrium. Experimental measurements and dimensional analysis show the relationship between entropy and energy as

(2.1)

$$\frac{\partial S}{\partial U} = constant = \left[\frac{J/_K}{J}\right] = \left[\frac{1}{K}\right] = \frac{1}{T}$$

Leading to the expression

(2.2)

$$\partial U = \partial Q = T * \partial S \rightarrow TS = Q$$

Work can come in many forms, including mechanical (pressure and volume: PV), chemical (molecule number and energy: Nμ), ect. Thus, energy of a system consists of several different energy components and the first law of thermodynamics is typically rewritten as

(2.3)

$$U = TS + PV + uN$$

Thermodynamic parameters are considered in two classes: extensive (depends on system size: U, S, V, N) and intensive (does not depend on system size: T, P, u). An extensive and intensive parameter is multiplied to give a specific type of work, such as mechanical work created by a change in pressure and/or volume. However, this approach only makes sense from a macroscopic perspective.

Microscopic aspects of the system are explained with statistical mechanics. Statistical mechanics connects microscopic properties with macroscopic properties. If the individual molecules within the system are viewed, it is seen that each molecule has its own properties, such as energy and location. The specific ordering and distribution of individual component's properties is referred to as a microstate. There is a microstate for any possible configuration that the system may possess. Each microstate is a probabilistic

state for the system. However, many microstates will have the same bulk properties. Therefore, they will be grouped into a macrostate. The number of microstates per macrostate is referred to as the multiplicity of that macrostate. The most probable macrostate for a system is the one with the highest multiplicity. Thus, the most probable macrostate gives the system's bulk, ensemble properties.

Entropy is the thermodynamic parameter which bridges macrostates and microstates. Consider two isolated systems in thermal contact with each other. The number of molecules and volume of each system is fixed. For simplicity, the molecules are non-interactive. Therefore, no work can be done on the system, and the energy of the system is only dependent on the heat exchange which affects temperature and entropy. The total energy of the two combined systems is given as the sum of the individual systems.

(2.4)

$$U_{tot} = U_1 + U_2$$

Here U_{tot} is the total energy of the system and U_i is the energy inside system i. Due to the isolated manner of the system, the total energy of the system remains constant as long as no external energy is added. Each microstate for the system is dependent upon the energy in the system such that $\Omega_i(U_i)$, where Ω_i is the i-th microstate for a system with energy U_i.

Since probability is multiplicative, the number of microstates for the entire system is given as

(2.5)

$$\Omega_{tot}(U_{tot}) = \Omega_1(U_1) * \Omega_2(U_2)$$

This expression can be rewritten to depend only upon the total energy and the energy of one system, using the total energy relationship.

(2.6)

$$\Omega_{tot}(U_{tot}) = \Omega_1(U_1) * \Omega_2(U_{tot} - U_1)$$

The function dependencies show that the value of total microstates is dependent upon the total energy and the energy in one system. At equilibrium, the system is in the highest probable state and thus, the microstates are maximized. This creates a relationship where the total derivative of the microstates with respect to energy will be zero, resulting in

(2.7)

$$\frac{d\Omega_{tot}}{dU_{tot}} = \frac{\partial\Omega_1}{\partial U_1} * \Omega_2 + \Omega_1 * \frac{\partial\Omega_2}{\partial U_2}\frac{\partial U_2}{\partial U_1} = 0$$

Knowing the change of the energy for one system with respect to the other will simply give negative one, rearrangement gives

(2.8)

$$\frac{1}{\Omega_1} * \frac{\partial\Omega_1}{\partial U_1} = \frac{1}{\Omega_2} * \frac{\partial\Omega_2}{\partial U_2}$$

Leading directly to a logarithmic function.

(2.9)

$$\frac{\partial \ln(\Omega_1)}{\partial U_1} = \frac{\partial \ln(\Omega_2)}{\partial U_2}$$

Evaluated at equilibrium, each side of the equal sign will give a constant, such that

(2.10)

$$\beta = \left(\frac{\partial \ln(\Omega)}{\partial U}\right)$$

This relationship, in combination with the thermodynamic relationship $\left(\frac{\partial S}{\partial U}\right)_{NV} = \frac{1}{T}$ results in

(2.11)

$$\frac{\partial S}{\partial U} * \frac{\partial U}{\partial \ln(\Omega)} = \frac{\Delta S}{\Delta \ln(\Omega)} = \frac{1}{\beta * T}$$

Where β becomes Boltzmann's constant.

It was first Boltzmann who had the thought of connecting statistical mechanics and thermodynamics, but Planck was the one to explicitly write the formula well known entropy equation.

(2.12)

$$S = k \ln(\Omega)$$

Thus, the second law of thermodynamics states a system will maximize its entropy, the equivalent to a system occupying the most probable state. However, in nature the energy is always trying to minimize itself. Therefore, equilibrium is a balance between energy and entropy terms.

Thermodynamic Potentials

Conservation of energy states energy cannot be created or destroyed, but there are various manners in which energy can take shape. Energy can flow between systems in thermal contact or particles can flow between systems with permeable membranes. It is

helpful to use energy potential functions which allow a mathematical function of state variables to describe the potential energy accessible to the system. There are various situations which can be measured; thus, there are various types of energy potentials depending upon the limits on the measured system.

The most basic thermodynamic potential is internal energy (U). Internal energy is used for isolated systems (closed, isochoric, adiabatic) where energy cannot be transferred in or out. Thus, the total energy change for isolated systems will always be zero ($dU = 0$). However, most systems are not isolated. For situations, energy has different limitations or constrictions. For example, a heat bath allows energy transfer to a system via heat. In a closed isothermal (constant temperature) system, heat is added via thermal contact with the heat bath. A Legendre transform on internal energy transforms it to Helmholtz free energy. Legendre transformations can be performed for addition constraints to obtain the thermodynamic potential that relates to the system of measure (Figure 2.1, Table 2.1). It is crucial to ensure that the energy potential matches the physical limitations of the system.

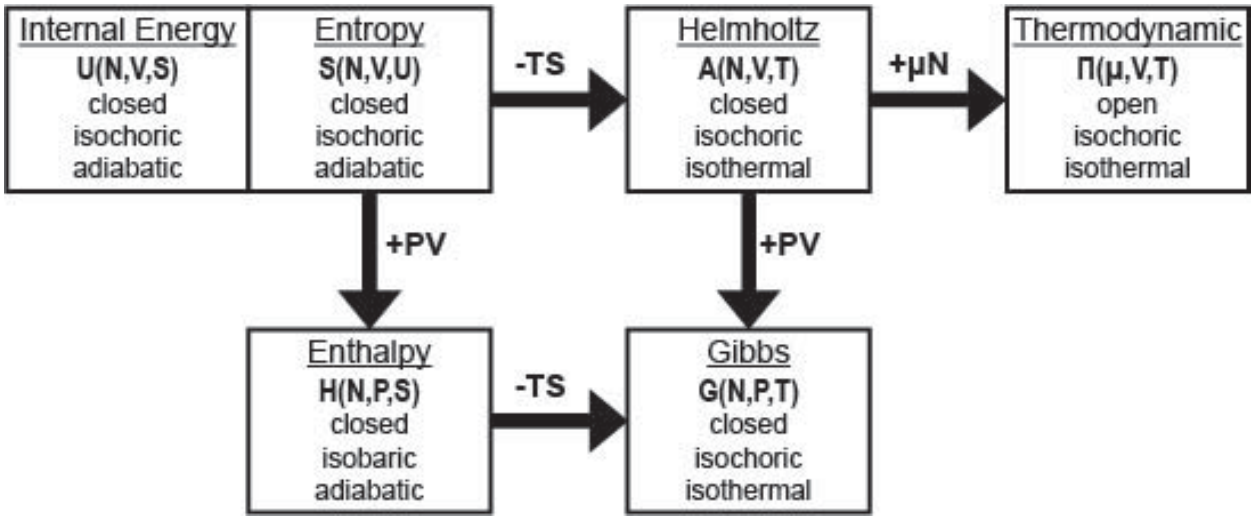


Figure 2.1 **Flow chart explaining Legendre Transforms for each thermodynamic potential** depending upon the system in question's set-up and properties.

Thermodynamic Potential	System Conditions	Equation	Derivative
Internal Energy U(N,V,S)	Closed Isochoric Adiabatic	$U(N,V,S) = TS - PV + \mu N$	$dU = T\,dS - P\,dV + \mu\,dN$
Entropy S(N,V,U)	Closed Isochoric Adiabatic	$S(N,V,U) = \frac{U}{T} + \frac{PV}{T} - \frac{\mu N}{T}$	$dS = \frac{1}{T}dU + \frac{P}{T}dV - \frac{u}{T}dN$
Helmholtz Free Energy A(N, V, T)	Closed Isochoric Isothermal	$A(N,V,T) = -PV + \mu N$	$dA = -P\,dV + \mu\,dN$
Thermodynamic Potential Π(u,V,T)	Open Isochoric Isothermal	$\Pi(u,V,T) = -A + uN = PV$	$d\Pi = S\,dT + P\,dV + N\,du$
Gibbs Free Energy G(N,P,T)	Closed, Isobaric, Isothermal	$G(N,P,T) = A + PV = U - TS + PV = uN$	$dG = -S\,dT + V\,dP$
Enthalpy H(N,P,S)	Closed Isobaric Adiabatic	$H(N,P,S) = U + PV$	$dH = T\,dS + V\,dP + u\,dN$

Table 2.1 **Common Thermodynamic potentials with the correct system conditions, equation, and derivative.**

Gibb's Mixing Paradox

Imagine a box full of gas molecules. The box contains a certain number (N) of molecules that cannot escape (closed system). Further, the box's walls are flexible, allowing it to adjust volume as needed to keep a constant pressure (isobaric). Last, the box is in thermal equilibrium with the air surrounding it (isothermal). For such a situation, Gibbs Free Energy is the thermodynamic potential required. The box system will contain an entropy (S) and volume (V) dependent upon the controllable parameters N, temperature (T) and pressure (P). For this example, the gas is made up of ideal molecules with no space or interactions. The entropy for an ideal gas is given by

(2.13)

$$S = Nk \ln V + \frac{3}{2} Nk \left\{1 + \ln\left(\frac{2\pi mkT}{h^2}\right)\right\}$$

If another box filled with the same molecules at the same number density is introduced in thermal contact with the first box, the total extensive thermodynamic parameters will increase to be the sum of both boxes combined.

(2.14)

$$S_{tot} = S_1 + S_2$$
$$U_{tot} = U_1 + U_2$$
$$V_{tot} = V_1 + V_2$$
$$N_{tot} = N_1 + N_2$$

While the intensive thermodynamic parameters remain unchanged

(2.15)

$$T_{tot} = T_1 = T_2$$
$$P_{tot} = P_1 = P_2$$
$$\mu_{tot} = \mu_1 = \mu_2$$

When the dividing wall between the two boxes is removed, the molecules are free to move between the two boxes and mix. Since entropy is dependent upon $\ln(V)$ it is no longer a strictly extensive parameter leading to a change of entropy associated with mixing

(2.16)
$$\Delta S_{mix} = S_{new} - (S_1 + S_2) = k\left[N_1 \ln\left(\frac{V_1 + V_2}{V_1}\right) + N_2 \ln\left(\frac{V_1 + V_2}{V_2}\right)\right] > 0$$

Since the change of entropy is positive, the mixing process must be irreversible.

However, when all molecules are identical, the divider can be reinserted, separating the systems back into the original configuration. In this case, $\Delta S_{mix} = 0$. Using the microstates to determine the change of entropy

(2.17)
$$\Delta S_{mix} = S_{new} - (S_1 + S_2) = k[\ln\{(N_1 + N_2)!\} - \ln(N_1!) - \ln(N_2!)]$$

This expression can be modified with a simple ad hoc term $k * \ln(N!)$. Hence, for mixing systems of the same molecule with the same temperature, the change in entropy becomes zero, fixing the Gibbs Paradox of mixing.

(2.18)
$$\Delta S_{mix} = k\left[(N_1 + N_2)\ln\left(\frac{V_1 + V_2}{N_1 + N_2}\right) - N_1 \ln\left(\frac{V_1}{N_2}\right) - N_2 \ln\left(\frac{V_2}{N_2}\right)\right] = 0$$

The previous situation is a perfect example of non-interacting ideal molecules. Such situations rarely occur in real-life situations. Thus, mixing may alter the energy within a system, particularly with interacting molecules. Therefore, mixing becomes a competition/balance between maximizing the entropy and minimizing the energy.

Phase Separation

Even a small impurity will drastically increase the entropy when mixing two different molecule types. This will drive systems into homogenous mixing. However, there is energy associated with molecules interacting with each other. If a high enough energy cost is associated with interacting, the system will be driven away from mixing and remain in two phases. This shows mixing is a balance between energy and entropy. When the energy cost is higher than the entropy contribution, phase separation will occur.

Experimental systems which study mixing typically occur in closed, isobaric, isothermal conditions leading to the use of Gibbs Free Energy ($G = H - TS$) to describe the system. Since the pressure and number of molecules in the system remain constant, the enthalpy (H) of the system will remain constant with the entropy (S) scaling with inverse temperature. Therefore, mixing becomes a balance of enthalpic and entropic forces. When mixing a solution of molecules, A and B, contamination of molecules will increase the entropy, but if there is an energy cost associated with A-B interactions, mixing will increase the energy. When the energy cost is too high to overcome the entropy, the Gibbs free energy curve for the system will appear with two local minima resulting in two stable equilibria for two different relative concentration values between A and B resulting in two stable phases (Figure 2.3). Since entropy is scaled by temperature, there will be a temperature at which the entropic considerations outweigh those of the energetic barrier for interaction. This point is referred to as the critical temperature. Above the critical temperature, no phase separation occurs.

Typically, phase separation is explained with a phase diagram (Figure 2.2). For phase diagrams the concentration of one component is listed upon the x-axis. The y-axis represents some environmental factor, typically temperature, though it can also be pH, salt concentration, etc. depending upon the system. A coexistence curve is then drawn to show conditions in which the two phases coexist, with any condition encompassed by the curve separating into two phases while any condition outside the curve remains in one phase. Experimentally phase diagrams are generated by assays where the concentration is constant while varying the environment, and vice-versa. Experimenters place discrete yes/no marks to build the coexistence curve. While such methods are beneficial to classify mixing behavior, further modeling is required to fully understand the mixing behavior.

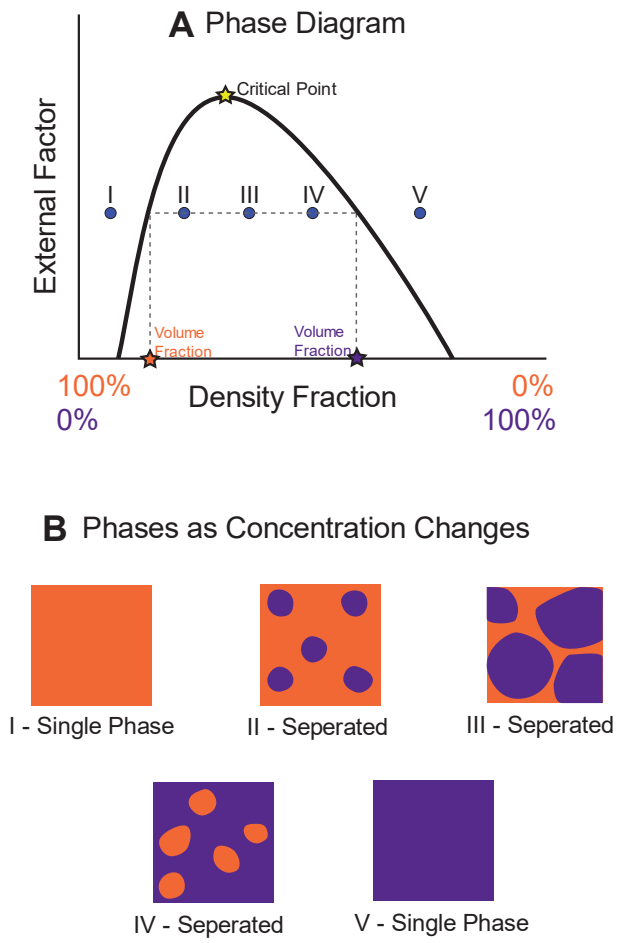


Figure 2.2 **Visual Description of Phase separation.** A) Typical phase diagram is shown in black. Any external factor could go on the y-axis, such as temperature or salt concentration. The x-axis shows the volume fraction of each molecule in the mixture, with the orange molecule shown in orange and the purple molecule shown in purple. When the mixing situation occurs outside of the curve (points I, V), the system will remain in a single phase. Within and under the curve (points II, II, IV) the system will phase separate. The volume fraction of each component within the phase separated fraction is found by the location of the curve along the desired external factor value (shown by the gray dashed line). The orange dominant phase will have the volume fraction according to the left of the curve, while the purple dominant phase is on the right, indicated by the stars on the x-axis in the respective color. The yellow star at the top of the curve indicates the critical point, above which phase separation does not occur. B) Visual sketch of mixture for each point I, II, III, IV, and V shown on part A. As the volume fraction of the purple molecule increases, the purple dense phase increases in volume until it becomes the bulk phase.

Flory-Huggins Theory

There are various models for phase separation. The most used for biological systems is Flory-Huggins theory for molecules and polymers. Flory-Huggins theory looks at the Gibbs Free Energy between polymer and solvent mixtures and interactions. Here the energy and entropy change are derived by modeling polymers and monomers as spheres that take up positions within a three-dimensional lattice (Figure 2.3). Since the size difference between the polymer and monomer is so large, a volume fraction is used to compare the relative "amount" of one to the other. The volume fraction is analogous to a mole fraction but is weighted to take account of the relative sizes of the molecules. This gives a weighted ratio regarding the sizes of the molecules where Φ_i is the volume fraction of species i, x_i is the number of monomer units in species i, N_i is the number of species i in the solution, N_{sites} is the total number of locations sites in the lattice which depends upon the volume of the mixture.

(2.19)

$$\Phi_i = \frac{x_i N_i}{N_{sites}}$$

When assuming each lattice site is occupied and the number of units for the solvent is one, the following relationship is obtained for the number of lattice sites. Where subscript p represents the protein/polymer and subscript s represents the solvent.

(2.20)

$$N_{sites} = \sum_i x_i N_i = x_p N_p + N_s$$

This results in the probability of a certain species occupying a space on the lattice equivalent to its volume fraction. This allows an expression for the change of entropy given the entropy and probability relationship. The change in entropy due to mixing, as modeled by a polymer random placement, is give as

(2.21)
$$\Delta S_{mix} = -k \sum_i N_i \ln(P_i) = -k \sum_i N_i \ln(\phi_i) = -k \left[N_p \ln\left(\frac{x_p N_p}{N_{sites}}\right) + N_m \ln\left(\frac{N_m}{N_{sites}}\right)\right]$$

However, Gibb's Free Energy is a combination of enthalpy and entropy considerations. There is an energy cost associated with molecules being in proximity and interacting with each other. In the Flory Huggins model, molecular interactions occur between lattice contact points. If the polymer is made up of the same monomer, three different interactions can occur.

(2.22)
$$\omega_{pp} = \text{polymer and polymer}$$
$$\omega_{ss} = \text{monomer and monomer}$$
$$\omega_{sp} = \text{polymer and monomer}$$

Each interaction may contain a different energy cost or benefit. The energy change associated with forming new contacts involves the loss of energy from breaking the initial contacts combined with energy gained due to new contacts forming. For a single change from same contacts to mixed contacts, the energy change is given as

(2.23)
$$\Delta\omega = \Omega_{sp} - \frac{1}{2}\left(\Omega_{pp} + \Omega_{ss}\right)$$

When the change in energy is negative mixing is preferred due to the decrease in enthalpy while increasing entropy, leading to homogenous mixing. When the energy change is positive, there is a competition between enthalpy and entropy. When the enthalpic term dominates, molecules will associate with similar molecules resulting in phase separation.

These results show the energy change of mixing dictates the number and composition of phases in the mixture. Thus, the energy change for the entire system needs to be determined. This involves calculation of the total number of mixed contact points multiplied by the energy change per contact. The total number of polymer contact points can be found as the total number of polymer units ($x_p N_p$) times the number of neighbors (z). Then the probability of finding a solvent as a neighbor, weighing the total polymer contacts by the solvent volume fraction, resulting in the total number of polymer-monomer contacts. The total number of contact points will be equal for polymer-monomer and monomer-polymer.

(2.24)

$$z(x_p N_p)\phi_s = zN_s\phi_p$$

The total energy of the system becomes the number of mixed contact points times the change of energy per contact. The polymer-solvent interaction parameter (χ_{sp}) is introduced to simplify this equation by combining the change in energy and number of neighbors, which are both system dependent.

(2.25)

$$\chi_{sp} = \left(\frac{1}{kT}\right) z * \Delta\omega$$

$$\Delta H = (\Delta\omega * z) * \left(N_s \phi_p\right) = \left(kT\chi_{sp}\right) * \left(N_s \phi_p\right)$$

Thus, the determining factor for phase separation is χ_{sp}, the interaction parameter. When χ_{sp} is very large, the energy cost for mixed interactions is large, driving to phase separation. When χ_{sp} is small or negative, mixing decreases the enthalpy driving a homogenous mixing (Figure 2.3).

Bring everything together, the total change in Gibbs free energy due to mixing becomes an expression combining entropy and enthalpy terms. Here N, number of molecules, is converted to n, moles of molecules, where Avogadro's number is absorbed into Boltzmann constant to give the ideal gas constant, R.

(2.26)

$$\Delta G = RT\left[n_s \phi_p \chi_{sp} + n_s \ln(\phi_s) + n_p \ln\left(\phi_p\right)\right]$$

When the interaction parameter is less than one, the entropy term dominates, and the system is completely mixed. Since natural log of a number less than one will always be negative. When the interaction parameter is large, the enthalpy term dominates pushing the system to phase separate.

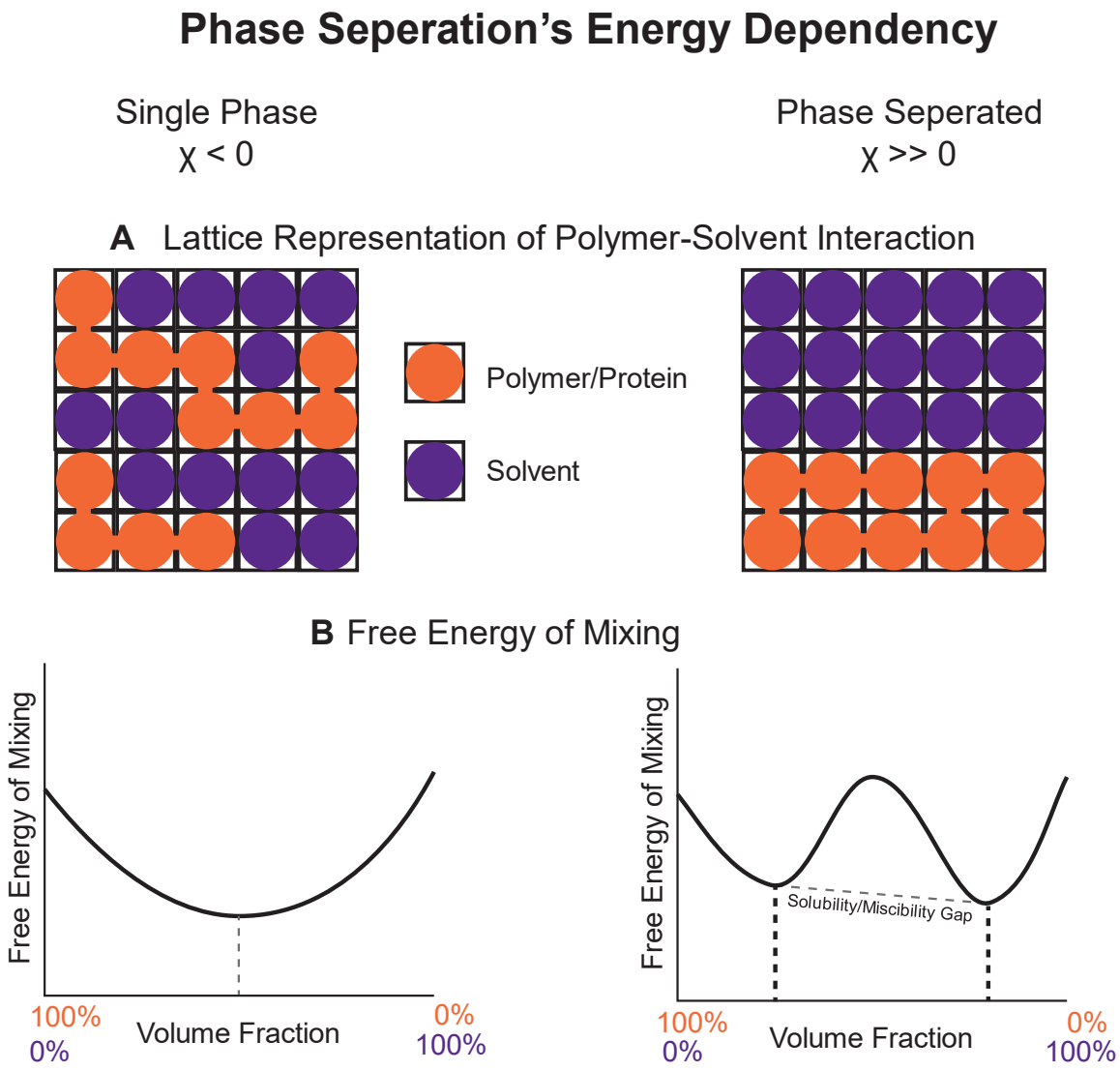


Figure 2.3 **Lattice layout and energy diagrams for mixing.** When in the interaction parameter, χ, is negative, a homogenous mixture will occur, as seen on the left. When χ is large, to phases occur within the system, as seen on the right. A) Lattice occupancy for a single phase (left) and two phases (right) system. B) Free energy of mixing for a single phase (left) and two phase (right) system.

CHAPTER THREE

SIMULATING BIOMOLECULAR CONDENSTATES

Determining the Interaction Parameter (χ)

For Flory-Huggins theory, the most important driving forces for phase separation is the interaction parameter (χ) between the polymer and the solvent. The interaction parameter (χ) is found experimentally to be linearly dependent upon inverse temperature (174,175), consistent with (eqn. (2.25)). Flory-Huggins theory has been used to understand phase separation and crystal formation between pharmaceutical products (176) including how to optimize drug load delivery using nanocarriers (177). Experimentally, the interaction parameter can be found through annealing polymers (178). However, obtaining the interaction parameter through computational methods has been more difficult (179-181). Since phase separation simulations requires multiple interacting polymers, full atom simulations are too computationally expensive leading to course graining models (182). Force fields for coarse grained systems require optimization for the system, therefore specific force fields are used solely to determine phase separation for biomolecular condensation formation (183). Fortunately, there have been great advances in coarse graining and running multiple simulations at various temperatures is manageable through use of LAMMPS (184).

Modifying eqn. (2.26), the free energy of mixing in terms of the volume fraction of the polymer or protein is given as

(3.1)

$$\frac{\Delta G_{mix}(\phi)}{kT} = \frac{\phi}{N_1}\ln(\phi) + \frac{1-\phi}{N_2}\ln(1-\phi) + \chi\phi(1-\phi)$$

Where N_1 and N_2 are the number of lattice spaces for the protein and water respectively.

At equilibrium, the first derivative of free energy must be zero for the system.

(3.2)

$$\frac{\partial \Delta G_{mix}}{\partial \phi} = \frac{1+\ln(\phi)}{N_1} - \frac{1+\ln(1-\phi)}{N_2} + \chi(1-2\phi) = 0$$

Further, the system will only phase separate if there are two separate local minima, thus the function must contain a convex portion, a location where the second derivative is negative, for phase separation to occur. If equation (3.3) is positive for all volume fractions, phase separation will not occur.

(3.3)

$$\frac{\partial^2 \Delta G_{mix}}{\partial^2 \phi} = \frac{1}{N_1\phi} + \frac{1}{N_2(1-\phi)} - 2\chi$$

Experimentally it is known that Chi should be linear with respect to inverse temperature and rearranging eqn. (3.2) for the interaction parameters give a means to extract the interaction parameter with only the volume fraction for each phase.

(3.4)

$$\chi = A + \frac{B}{T} = \left(\frac{1+\ln(1-\phi)}{N_2} - \frac{1+\ln(\phi)}{N_1}\right) * \left(\frac{1}{1-2\phi}\right)$$

Using the simulated phase densities to determine the interaction parameter, a linear fit can give values for parameters A and B, allowing the interaction parameter for the system to be determined. The interaction parameter will give the free energy of mixing as a

function of temperature. The free energy of mixing minimization for both volume fractions will result in the density of each phase when phase separated. This allows the phase diagram to be constructed as a function of temperature with only the volume fraction of high- and low-density phases at several temperatures, values obtainable with the appropriate computational approaches.

Coarse Graining Methods

Computational simulations are limited by computational resources and run time. A high degree of freedom for a system is the largest contributor affecting the large number of computational resources and long run time required. Classical all atom simulations tend to be limited by a tradeoff between the size of the system and the length of the simulation. One way to limit the degree of freedom in a simulation is to combine atoms into one group or bead. Such a method, referred to as coarse graining, increases computational efficiency by decreasing the degree of freedom in the system (185). Coarse graining can drastically increase the time scale simulations can achieve, obtaining up to second time scales (186). The increase in achievable timescale allows computational methods to efficiently match the same time scale (microsecond or millisecond) as experimental spectroscopic techniques (187) and probe longer time scales of protein aggregation, allowing better insight into the entire aggregation process (188). Coarse graining can also increase the size of the system under investigation such as DNA and chromatin interactions which are too large to be probed with atomic methods

(189,190). Coarse graining large systems leads to simulations results that agree very well with experimental results regarding melting curves and persistence lengths (189).

Typical DNA coarse graining models involved dividing each base pair into three parts (phosphate, sugar, and base) represented by three oval 'beads' (190). Proteins can be coarse grained where each amino acid is represented by a single bead (191). There are many different methods of mapping between coarse grain models and full atomistic models (192). These methods drastically lower the degree of freedom for the systems, while giving accurate results that match well with experimental results. However, one must be careful not to eliminate too many degrees of freedom or important interaction sites.

With every simulation, the computational results are only as good as the force field for the system (193). Comparison of atomic and coarse grain models are being used to improve force field parameters for coarse grain simulation, leading to better consistency between coarse grain and atomistic models (194). The results is using atomic force fields to create coarse grain force fields. Some force fields are experimentally determined, common for coarse graining where a single bead represents an amino acid. In such cases, each amino acid has specialized interaction parameters which are used for a Lennard-Jones (LJ) potential. These force fields can accurately predict protein binding and interactions (85,191) as well as IDP, including Q-rich regions, behaviors (195).

However, there are draw backs regarding coarse grain models. In general, coarse graining smooths the energy landscape, leading to faster dynamics (196). However, other thermodynamic properties such as temperature, pressure, and density agree with atomic

models across various ranges (197). There is active development for out of equilibrium coarse grain models (198) and mixed models that combine coarse graining with certain fully atomistic regions of the system (108). It is important to understand the aim of the simulation to ensure the appropriate experimental design required for the simulation.

The advances in coarse graining have created a decent method for quickly verifying hypothesis before putting resources into a full-atomic simulation or an experimental set up (199). Combining coarse graining parallel computing, computational time drastically decreases allowing for larger systems or longer simulations (200). Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) is a commonly used simulation tool specifically designed for parallel computing of course grained approaches (Chapter 3.2) (201). Patches have been introduced to LAMMPS for a coarse grain Lennard-Jones-9-6 potential, making coarse graining simulations easier to run for coarse graining proteins by amino acids (202). For the simulations within this dissertation, an experimental, physic-based force field (203) is used following the set up and parameters from Dignon et al (184). Where the protein sections are coarse grained using a top-down model (204), where each amino acid is represented by a single bead.

Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)

Even with coarse graining techniques, running simulations large enough to predict and understand biomolecular condensation formation is computationally heavy, requiring large amounts of resources with long simulation times for a single simulation. To build a phase diagram, several phase separation simulations must be completed at multiple

temperatures to properly obtain the varying high- and low-density phase concentration temperature dependence. Thus, for a single sample, several simulations must be completed. Parallel computing decreases computational time by dividing the particles in the simulation to different processors which communicate with each other during various times steps within the simulation leading to better computational efficiency (200,205). Due to the ease of use and widely available scripts, LAMMPS is the most used open-source parallel simulator environments used for molecular dynamic (206). LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) has been filling the need for an easy-to-use parallel computing molecular dynamic simulation since it became open source in 2004 (201). It is used across various fields from material science to biological systems.

LAMMPS was designed to fill the need for easy access to simulation techniques for scientists who might not have a deep understanding of computational skills (207). With its versatile applications across temporal and spacial skills, it is widely used in material science and biological simulations. For users with little computational experience, LAMMPS scripts can be modified using the input file. Since the input file is read one line at a time, multiple simulation stages can easily be applied within a single input file. Commands within the input file include defining interaction parameters for atoms, boundary conditions and processor distribution. Input files can also include specific calculations and output file formats. Further, python scripts can be implemented directly into an input file for more complex calculations. More experienced computer coders can modify LAMMPS to run more specific scripts including introduction of user

defined pairwise interactions through the input file. This allows the user to define the interaction parameters to be used in the force field allowing various coarse graining force fields from solid state, soft matter, and biological systems dependent upon the system being probed by the user. More advanced users can create and develop LAMMPS sub-directories, referred to as packages, that can be used as a LAMMPS library to allow for more advanced modifications to simulations. LAMMPS libraries can be downloaded and installed for any user allowing for easy access and sharing of computational design for similar systems. LAMMPS libraries are compatible with python coding, allowing users to modify the LAMMPS libraries for their specific requirements using python.

LAMMPS is highly modifiable through modification of different, easy to change, parameters. The ease of modification allowed within LAMMPS allows for various systems and properties to be probed. Bond information between particles can be altered throughout simulations allowing the changing of bond information, such as bond breakage, formation, and type changes, to be explored (208). Further user modifications can alter the way energy is introduced into the system allowing for more realistic temperature controls for heating and cooling (202). The application of forces can also be modified, allowing for a better representation of the system being probed (209). The LAMMPS environment is even capable of exploring nonequilibrium (210,211) and multiple scale (212,213). However, some aspects remain constant regardless of the system of interest.

LAMMPS is largely controlled by fix commands. Fix commands can be inserted into the input file to define certain interactions or parameters which can change

throughout the simulation. Fix commands give easy methods of modifying LAMMPS simulations to fit various situations. When starting a simulation, the simulation box must be defined. There are various ways to define a simulation box including the shape of the box and boundary conditions. The simulation box can be shaped with orthogonal or triclinic dimensions. Boundary conditions may be periodic or non-periodic with fixed or shrink-wrapped boundaries. It is important to understand the system and what boundary conditions will best fit the simulation. Simulation box and boundary conditions can be modified throughout the simulation through a fix command within the input file.

Once the boundary conditions are set, the simulation must be split into different regions, referred to as partitions, that will be assigned to different processors allowing for the parallel computing aspect. The simulation box is spatially partitioned into non-overlapping subdomains. The boundaries of partitions must be parallel with simulation box boundaries. The default partition will break the simulation box into equal boxes. The default method is useful for simulations containing a uniform density. However, there are many situations where the density may not be uniform throughout the simulation space. Variable sized subdomains assign the same number of particles to each processor, resulting in uneven spacial splitting between processors. Dynamic load-balancing algorithms, such as recursive coordinate bisectioning (RCB), can help increase computational efficiency, since the efficiency of parallel computing is dependent upon the heaviest load upon a single processor (214,215). Partitions share a region of overlap by a set cutoff to ensure clean transition of particles between partition spaces. The particles within a partition are owned by that partition while particles existing within the

overlap are referred to as ghost particles. Each processor stores the information regarding atoms which it owns as well as ghost atoms. As the simulation runs, particles may transition between partitions. When a particle diffuses out of a partition and transitions from being owned to a ghost particle, a forward communication occurs within the processor. As a particle diffuses into a partition it goes from being a ghost particle to an owned particle, a reverse communication occurs in the processor.

CHAPTER FOUR

FLUORESCENCE BASED EXPERIMENTATION

Fluorophore Selection

There exist certain molecules that appear to 'glow' under certain conditions. Such molecules are referred to under many names including fluorophore, chromophore, and fluorescent dye. Naming convention tends to rely on structure of the molecules with small, synthesized organic molecules commonly referred to as fluorescent dyes with larger, biologically synthesized proteins referred to as fluorescent proteins. For the rest of the dissertation, such molecules will be referred to as fluorophores to encompass all such molecules.

Developing fluorophores with different physical properties is a large, active research topic (216). There are several physical properties that are used to classify fluorophores including absorption and emission range, excitation lifetime, quantum yield, photostability, and environmental influence (217). To fully understand the importance of these properties, one must first understand the physical phenomena that leads to the 'glowing' effect. Fluorophores contain a chemical structure with a ground state energy (S_o) and a higher-energy excited state (S_1). The energy barrier between the ground state and excited state depends upon the overall chemical structure of the fluorophore. Some fluorophores contain several different excited states, accessible by different energy barriers. Commonly, fluorophores absorb an external energy source, typically a photon, which contains the same energy as the energy barrier, which transitions the electrons to

the excited state. For simplicity, it will be assumed that the fluorophore is being excited by a single photon, though other excitation manners exist (218-220). The fluorophore remains in the excited state for a certain time, referred to as the lifetime of the excited state, or simply the lifetime, after which the fluorophore relaxes back into the ground state and released the energy back into the environment. Fluorophores are designed to optimize the released energy into the form of a photon as radiative emission. This is what is seen as the 'glowing' of the fluorophore. However, there are several mechanisms in which the fluorophore can relax back to the ground state which are non-radiative, energy released in a form other than a photon. It is important that fluorophores function as designed and there exist several parameters used to describe fluorophore functionality.

The largest difference between fluorophores is their absorbance and emission spectra (221). Since fluorophores are constructed from complex atomic structures, their energy bands consist of very close vibrational energy levels. This creates a range of energy values which can be absorbed by the fluorophore. Assuming the fluorophore will be excited by a photon, energy can be replaced by wavelength. Experimentally, the absorbance of different wavelengths can be determined for each fluorophore building an absorbance spectrum. An absorbance spectrum will have a peak wavelength where the fluorophore optimizes absorbance with less absorbance in the neighboring wavelengths creating a curve of absorbed wavelengths due to the fluctuations of energy levels due to vibrational energy. Few fluorophores will contain multiple peaks depending on the chemical structure and energy levels. Once the photon is absorbed, the fluorophore becomes excited. Energy is lost in the dissipation of vibrational energy, resulting in a

longer wavelength, lower energy photon being emitted. The emitted photons, referred to as emission, also result in a range with a peak showing the most occurring wavelength with counts decreasing along the neighboring wavelengths. Collecting the counts for each wavelength creates an emission spectrum for the fluorophore. The peak locations of the absorption and emission spectra are called the maximum absorption or excitation and maximum emission, respectively. Selecting fluorophores with the appropriate absorption and emission spectra is very important. Most biological samples naturally contain some color or fluorescent that can increase the background making the detection of fluorophores with similar spectra properties to the background difficult to measure (222). For example, when designing and experimenting using smooth muscle cells, which have a purple/red tint, one should avoid fluorophores with similar emission spectra wavelengths such as mCherry and pick a fluorophore that appears closer to green such as GFP.

Since fluorophores do not absorb all wavelengths of light equally, the molar absorption coefficient, also referred to as the extinction coefficient, gives a measure of how well a certain fluorophore absorbs a specific wavelength. Since absorption depends upon the number of molecules within the light path, the molar absorption coefficient has units of inverse Molarity times centimeter and is most often used within Beer–Lambert's law where log of the ratio of intensity of light put into the system by the light that passes through is the absorbance and is equal to the concentration of molecules, the pathlength, and the molar absorption coefficient. It is important for fluorophores to have a large molar absorption coefficient to ensure that the fluorophore has a high probability of

excitation when encountering a photon. Further, many research setups have a limited number of lasers with specified wavelengths. It is important to ensure that available lasers are compatible with the selected fluorophore's wavelength associated with a higher molar absorption coefficient.

Not all absorbed photons are released back as photons. This loss of photon counts varies on several factors and is parameterized as the quantum yield, the ratio between the number of photons emitted by the number of photons absorbed. Fluorophore researchers strive to optimize the quantum yield by maximizing it to get it as close to one as possible. The quantum yield depends upon several environmental factors including pH, salt concentration, polarity of solution, and base solvent. It is important to consider the environment when designing experiments. Multiplying the quantum yield and molar absorption coefficient gives the brightness of the fluorophore. The brightness is a good manner of understanding the emission process efficiency for a fluorophore at a specified wavelength. Often fluorophores are selected based on the experimental environment, since biological samples require specific conditions. The photostability, chemical influence and solubility of the fluorophore in response to the experimental environment is highly important to consider when selecting which fluorophores to use in an experiment. A well selected fluorophore will be easily visible and detected throughout the experiment.

One phenomenon that can impair the brightness of a fluorophore is quenching, a mechanism that requires the interaction between the fluorophore and a quencher (223). The quencher is a secondary molecule within the environment that receives energy

transferred from the fluorophore, placing the fluorophore back into the ground state without the release of a photon. Common quenchers include oxygen, aromatic rings, heavy metals, and even certain amino acids. Quenching comes in two forms: dynamic and static. Dynamic quenching occurs when the fluorophore diffuses into or collides with the quencher. Since dynamic quenching is dependent upon collisions and diffusions, it has a dependency upon time thus alters the average lifetime of the fluorophore. Static quenching occurs when the fluorophore binds with the quencher forming a complex which consistently releases energy in a form different from the expected photon. Static quenching does not change the lifetime of the fluorophore since the formed complexes are expected to remain stable over time lengths longer than the lifetime of the fluorophore. Since dynamic quenching alters the lifetime of the fluorophore, while static quenching does not, experiments which obtain the lifetime of the fluorophore can separate dynamic and static quenching.

Another situation which lowers fluorophores efficiency is alterations within their chemical configuration. Some altered no longer allow for radiative emission, resulting in the fluorophore entering a dark state. While in a dark state, fluorophore cannot receive excitation energy. Some fluorophores can recover from the dark state stochastically, after a certain amount of time. This mechanism can be manipulated and useful for certain experimental set ups such as Fluorescence Recovery After Photobleaching (FRAP) or to optimize single molecule wide-field imaging techniques by lessening the fluorescent molecules in view by taking advantage of fluorophore blinking, a property of some fluorophores which stochastically fluctuate between normal, bright states and dark states.

Blinking can help limit the amount of fluorophore labeled molecules within a frame for better particle tracking or localization. For example, Stimulated Emission Depletion (STED) imaging is a super-resolution microscopy technique that uses the dark state to overcome the spacial resolution limit given by conventional optical microscopy.

Another feature of fluorophores is they do not remain stable forever. Many fluorophores have a measurement life on the order of milliseconds to second. Due to the life span of fluorophores, many fluorescent experiments are limited to probing a specific fluorophore for a short period of time. Recent developments in fluorogenic have been developed to try and overcome these limitations by introducing fluorophores which only access the bright state when bound to their target. Controlling the bound and unbound rates, such fluorophores can be developed with fast bound rates to ensure quick binding and unbound rates close to the time it takes for them to naturally enter a dark state. These mechanisms work to refresh fluorophores throughout the experiment allowing for the same target molecule to be tracked for longer periods of time. Introducing multiple binding locations on the same target molecules will help reduce the 'dark gaps' which occur between a fluorophore unbinding before a fresh fluorophore binds to replace it. Since binding is stochastic, work needs to be done to optimize the number of binding locations to ensure a consistent tracking of molecules throughout the duration of the experiment.

It is essential to consider all the above fluorophore properties when designing an experiment. Adjusting the system by introducing the best fluorophore will result in the best results. Often it is difficult to optimize for a perfect fluorophore and trade-offs must

be considered. Given the experimental setup and underlying hypothesis, careful consideration should be given to the selection of fluorophores.

Labeling Strategies

Fluorophores are connected to the target of interest in several manners depending upon the target and the fluorophore selected for the experiment. Common target molecules include DNA, RNA, and proteins. For this dissertation, the focus will remain on proteins.

Fluorescent proteins are formed from amino acids, the same way all proteins are created. These amino acids tend to follow a common secondary structure resulting in a barrel shape formed from beta sheets protecting an internal chemical structure containing an aromatic ring which results in the fluorescent properties of the protein. Since fluorescent proteins are themselves proteins, they tend to be nanometers in size, with Green Fluorescent Protein (GFP), one of the first and commonly used fluorescent proteins, having a radius of gyration around 2.5 nm (224). Further modification on GFP has resulted in a variety of different fluorescent proteins like enhanced GFP (eGFP) (225) with mutations leading to a variety of different absorbance and emission ranges such as yellow fluorescent proteins (YFP) and mCherry (226,227). Because fluorescent proteins are proteins, they can be tagged to a protein of interest by recombinant DNA methods. This involves splicing together two DNA sequences from different host organisms and introducing the new DNA sequence to a host system for expression. The result produces the target proteins with the fluorescent protein connect at the N- or C-terminal depending

upon the DNA plasmid design. Such plasmid constructs can be introduced into a cell line which benefits the manner of the experiment, being in vivo or in vitro.

On the other hand, synthesized organic molecules also known as fluorescent dyes, are much smaller than fluorescent proteins and tend to remain brighter and more stable over time. Fluorescent dyes are typically connected to a carbon chain linker with a reactive group on the end. This reactive group is specified to react with a certain chemical trait on the target molecule to ensure the molecule is tagged at a certain location. Such is the case for the GluN2B study in this dissertation, maleimide groups are used at the reactive group at the end of the carbon chain which reacts with sulfur within cystine amino acids. Thus, mutating the target protein to remove native cystines and introduce cystines at strategic labeling locations within the protein allows for the protein of interest to be targeted along its sequence instead of simply a tail end. Care must be taken when selecting mutation locations within the protein. The native function of the protein must be ensured, eliminating reactive domains along the protein or amino acids important for secondary structure (228). There are several tools currently in use to computationally predict appropriate labeling locations within proteins, including projects on going within the Single Molecule Biophysics Laboratory (SMBLab) here at Clemson.

While reactive groups work very well for in vitro experiments, there are many issues for in vivo experiments, largely including non-specific binding to any binding target within the host system. For in vivo experiments, immunofluorescence (IF), where the protein of interest is tagged with an antibody, becomes very useful. In IF, the fluorescent dye is attached to the end of an antibody where the antibody has a specific

targeting site which allows for more specified binding than more general reactive groups. However, antibody development can be difficult and intensive often resulting in a high background or non-specific binding. Depending upon the sensitivity of the experiment, directly tagged fluorescent proteins may be the best option for in vivo experiments.

Determining the best labeling method for fluorescent experiments is essential. The labeling location can alter the system or change the properties of the measurement. To ensure the results are expected control experiments, correct labeling locations and methods must be considered.

Fluorescence Detection Methods

When using fluorescence to probe a system, it is important to understand what physical phenomenon is being measured and reported to extrapolate the concept of interest. Fluorescent data, when used correctly, offers a wealth of information. However, the quality of the information gathered is limited by the available experimental equipment and various other experimental factors. The easiest fluorescent experiment to imagine is in vivo cell imaging where a protein of interest is tagged with a fluorophore to determine the proteins localization within the cell. In this case, widefield images, taken through an optical microscope where a laser is used to excite the attached fluorophore and a Complementary Metal Oxide Semiconductor (CMOS) camera is used to take images. Such experiments are highly beneficial in understanding general protein location within the cell while requiring simple, easy to access equipment. By comparing the intensity from each pixel that builds the image, with the assumption each protein of interest

contains a fluorophore label, and each fluorophore has a standardized brightness, number of the fluorophores in each location can be determined. Multiple lasers and fluorophores can be added to detect multiple different types of proteins of interest at once. Fluorescence Recovery After Photobleaching (FRAP) has a similar experimental set up but in this case, a powerful laser is used to photobleached, forcing the fluorophore into a dark state, within a small, controlled area. As fluorescently tagged proteins diffuse around the space, the rate of fluorescence recovery can be measured and covered into an average diffusion time for the fluorescently tagged protein ensemble. If the concentration of fluorophores is low enough, single molecule information can be gathered. The fluorescent signal from each fluorophore can be traced over time by comparing changes between frames. With the help of machine learning and increased frame rates, diffusion times and trajectories of individual molecules can be traced (229). This gives insight into individual protein behavior instead over averaging over an ensemble.

The given examples are powerful imaging techniques, but since CMOS cameras cannot capture certain light properties, such as polarization, there is a loss of fluorescent information that can be used to deeper probe into the physical system of the cell. Further limitations of CMOS cameras arise since the temporal resolution is limited by the capture rate of the camera, typically to the limit of 10s of milliseconds, though pixel binning can increase this speed, and the spacial resolution is limited by the wavelength of light and pixel size of the camera. Special care and clever ideas have led to the development of super-resolution microscopy, which allows imaging to advance past the standard spacial resolution limitations set by traditional optical microscopy methods (230-237).

With extra finances, more comprehensive experimental set ups can be constructed to detect additional fluorescent information such as the polarization of the emitted fluorescence. If a widefield image is not necessary to probe the desired physical property of the system of interest, confocal methods allow faster data accusations with photon-counting detectors such as avalanche photodiodes (APDs) and photon multiplier tubes (PMTs). Such photon-counting detectors measure photon arrival time with temporal resolution much faster than typical widefield cameras, often up to the picosecond range. This allows accurate measurements of fluorescence lifetimes which are often on the nanosecond time scale by detecting single photon arrival times and building a histogram of photon arrival time over the course of a measurement. This time scale can capture individual protein dynamics such as configuration changes and intrinsically disordered region movements.

Understanding the accessible tools and equipment available as well as the underlaying physics, temporal and spacial resolution required to probe a hypothesis are essential for experimental design of biological systems. When the experimental set up is deeply understood, experimental measurements can be properly optimized to ensure the best data is obtained allowing for the full accessible knowledge to be obtained from the dataset.

Fluorescence Correlation Spectroscopy (FCS)

Fluorescent Correlation Spectroscopy (FCS) is based upon time-dependent intensity fluctuations which result from dynamic processes. Very often this process

comes from translation diffusion into and out of a defined confocal volume created by focusing a laser beam through a confocal pinhole. The measurement confocal volume is altered by the pinhole size and wavelength of laser light being used, where a larger pinhole and longer wavelength leading to larger confocal volumes. Typical FCS experiments probe very small volumes on the order of a femtoliter.

As random diffusion drives fluorophores into and out of the confocal volume, time-dependent fluctuations occur within the measured fluorophore intensity of the measurement. The number of fluorophores within the confocal volume is directly related to the intensity of the measurement. However, intensity also depends upon the fluorophore location within the confocal volume. Fluorophores along the edges of the confocal volume have a lower intensity, appearing dimmer than fluorophores within the center of the confocal volume. Therefore molecular information can be obtained from FCS measurement by analyzing the rates and amplitudes of the intensity fluctuations using Poisson statistics (238).

To ensure measurements of only a single molecule, single molecule experiments are typically performed at low concentrations around picomolar of fluorescently tagged molecules. This low concentration, in combination with the small confocal volume, creates an average of approximately 0.6 molecules within the confocal volume at a time. This can be thought of as having a single molecule within the confocal volume around 50% of the time with very low probabilities of multiple molecules occurring within the confocal volume. These low concentration in combination with the desire to obtain information on fluctuations results in FCS measurements requiring some very specific

and advanced equipment to increase the signal to noise ratio ensuring measurement values are distinguishable from the background.

Intensity fluctuations often occur due to translation diffusion of fluorophores in and out of the confocal volume. As fluorophores diffuse into the confocal volume, they are excited by the laser and emit a photon. If the fluorophore remains within the confocal volume, it will be excited again, resulting in another photon emission. Thus, there is a burst of emitted photons due to multiple excitation-emission cycles from a single fluorophore dependent upon the dwell time of that fluorophore. If the fluorophore diffuses rapidly the dwell time is short leading to a short-lived photon burst time. If the fluorophore diffuses more slowly, the fluorophore remains in the confocal volume longer resulting in a longer photon burst time. This creates a relationship between the diffusion rate of the fluorophore and the dwell time within the confocal volume.

The autocorrelation function, G(τ), explains how intensity measurements fluctuate over time by comparing the fluorescence intensity at a time, t, to a delay time, τ.

(4.1)

$$G(\tau) = \frac{< F(t) \cdot F(t+\tau) >}{< F(t)^2 >}$$

The autocorrelation function will have dependencies upon the experimental set up, including the confocal volume geometry. The confocal volume is modeled to fit a 3D Gaussian with symmetry in the xy-plane where the radius of the xy-plane is given as ω_{xy} and the radius in the z-direction is denoted ω_z. For a system with fluctuations only due to simple translation diffusion, the autocorrelation function is given as

(4.2)

$$G(\tau) = G(0)\left(1 + \frac{4D}{\omega_{xy}^2}\right)^{-1}\left(1 + \frac{4D\tau}{\omega_z^2}\right)^{-\frac{1}{2}}$$

where G(0) relates to a τ equal to zero denoting no time delay, giving the average number of molecules being observed within the confocal volume. Since the translational diffusion coefficient, D, is directly related to size and geometry of the fluorophore and attached molecule, the diffusion time can allow one to understand the size of the molecule and any complexes that may be formed in the solution due to binding between two or more molecules.

Typical systems are rarely this well behaved and can contain intensity fluctuations due to ligand–macromolecule binding, rotational diffusion, internal macromolecule dynamics, intersystem crossing, and excited-state reactions. It is important to understand the physical and fluorescent properties of the measured system to ensure the correct fitting model is being applied. There are numerous different fitting models based on physical concepts allowing a large amount of information about the probed system to be obtained when proper analysis of fluctuations is applied. The fitting function will depend upon the photophysical properties of the selected fluorophore, local environment of the labeling location, and the biological system being probed. With the proper fitting function, FCS can offer a wide amount of information for various systems. For example, binding assays in which fluorophore properties will change upon binding giving reaction kinetics of the system.

Combining FCS with fluorescent lifetime information, a further wealth of information is gathered allowing multiparameter fluorescence detection (MFD) analysis

(239,240) resulting in the conformational dynamics of a system to be determined over a broad timescale (241-244).

Förster Resonance Energy Transfer (FRET)

Förster Resonance Energy Transfer (FRET) occurs when two fluorophores interact through weakly-coupled dipoles which oscillate upon excitation, resulting in a nonradiative, translation of energy from one fluorophore to the other. Such energy can only move 'downstream', from a fluorophore with a higher energy gap, referred to as the donor, to a fluorophore with a lower energy gap, referred to as the acceptor. The energy levels of the fluorophores must be compatible such that the emission energy of the donor overlaps with the excitation energy of the acceptor (Figure 4.1B).

Experimentally, FRET is measured by directly exciting the donor fluorophore, placing it within an excited energy state. The energy is then nonradiative transferred through an electric field which is generated by the oscillation of the dipole between the excited donor and ground state acceptor. This excites the ground state acceptor, placing it in a higher energy state. The acceptor can then decay back to the ground state through radiative or non-radiative methods (Figure 4.1A). Since FRET depends upon the dipole coupling of the fluorophores and resulting electric field, it can only occur within the near field zone of the electric field. In the near zone, 1 to 10 nanometers, the dipole is far enough away to ignore kinetic or collision effects, but close enough so that the electric field contribution can be ignored as well.

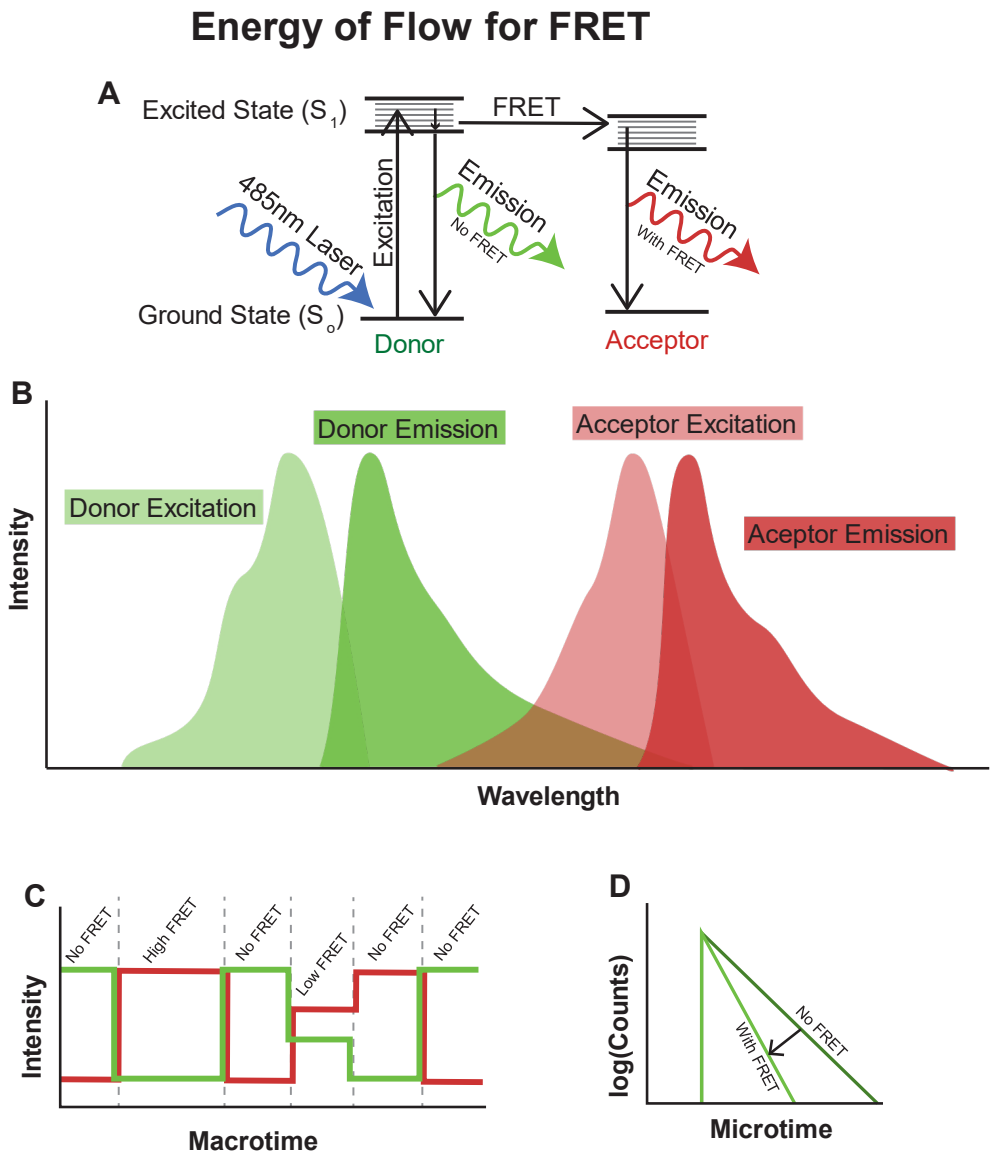


Figure 4.1 **Explanation of FRET phenomenon and data representation.** A) Jablonski diagram explaining the energy transfer during FRET. The donor is directly excited by a 485nm laser, shown in blue. When FRET does not occur, the donor emits a photon, shown in green. If FRET does occur, the acceptor emits a photon, shown in red. B) Spectra overlap between the donor emission and acceptor excitation must be present in order for FRET to occur. C) Common visual of intensity-based FRET experiment. Several excitation-emission cycles are averaged to obtain the fraction of donor and acceptor photons detected over the entire course of the experiment (macrotime). D) Common visual of lifetime-based FRET experiment. As FRET occurs, the lifetime of the donor is decreased. Lifetime histograms are formed from combining each excitation-emission cycle after the timing of the laser pulse (microtime).

FRET is a probabilistic phenomenon occur with a certain rate, k_T, depending upon the distance between the fluorophores. The rate of energy transfer is also referred to as the efficiency of energy transfer or FRET Efficiency, E. Measuring the FRET Efficiency allows for the distance between a donor and acceptor pair to be determined, typically with angstrom spacial resolution. Higher FRET Efficiencies correspond to closer distances, while a lower FRET Efficiency corresponds to further distances The FRET Efficiency and fluorophore distance, r_{DA}, follow an inverse sixth power relationship. Because FRET Efficiencies are dependent upon the system conditions for detection, it is important to measure standards with known distances for proper calibration. This can be done using DNA FRET standards which contain known donor-acceptor pair distances by using ridged DNA at known labeling sites to measure FRET (Figure 4.2). Since DNA is a rigid molecule, the distance between the donor and acceptor will remain constant for the entirety of the measurement and align along the static FRET line. The static FRET LINE is a given relationship between lifetime of the donor and FRET Efficiency that occurs when the donor-acceptor pair remain with a constant distance between each other and there are no dynamics (245).

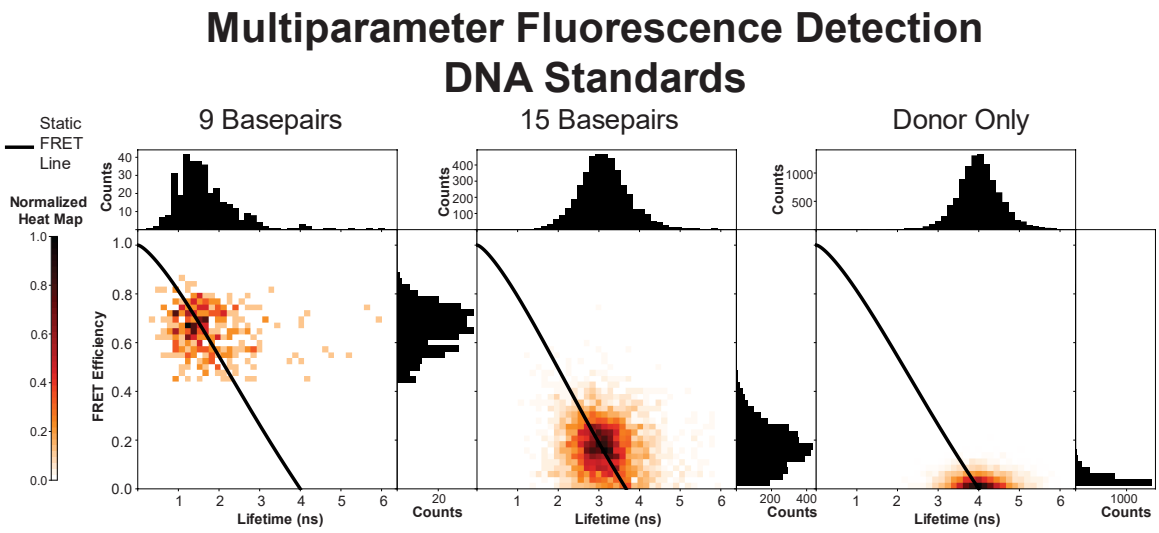


Figure 4.2 **Multiparameter Fluorescence Detection (MFD) data of DNA FRET Standards.** Control measurement used for system calibration for data analysis. The black line correlates to the static FRET line.

FRET Efficiency can be determined using intensity-based or lifetime-based measurements. Intensity based measurements require the fluorescence intensity from a sample with only the donor present, I_D, and fluorescence intensity of the donor in the presence of acceptor, I_{DA}. The donor intensity is used as a reference to refer to a situation in which no FRET occurs, giving a baseline for how the presence of acceptor leads to nonradiative energy transfer, which is assumed to be FRET based (Figure 4.1C). FRET Efficiency can also be found from the decrease in lifetime decay of the donor fluorophore leading in a lower donor lifetime (Figure 4.1D). In a similar manner to intensity-based FRET efficiency, the native donor lifetime is determined without acceptor present as a reference to determine the effect the acceptor has on the donor lifetime.

(4.3)
$$E = \frac{1}{1 + \frac{1}{k_T * \tau_D}} = \frac{1}{1 + \left(\frac{r_{DA}}{R_o}\right)^6} = 1 - \frac{I_{DA}}{I_D} = 1 - \frac{\tau_{DA}}{\tau_D}$$

FRET is a widely used phenomenon resulting in experiments to measure signaling, localization or probing as well as more complex experiments which use it for more accurate distance and dynamic measurements of proteins (246-251). There exist numerous commercial FRET fluorophore pairs. The magnitude of likelihood for FRET occurrence is based on the energy overlap and optical properties of the fluorophores. Thus, each FRET pair comes with their optimal measurement distances and Förster radius. The Förster radius, R_o, is the distance at the FRET Efficiency is 50% for the fluorophore pair. The Förster radius is based off independently measured parameters determined by $R_o^6 = \frac{9(\ln 10)\kappa^2 \Phi_D J^\lambda}{128\pi^5 n^4 N_A}$, where Φ_D is the quantum yield of donor fluorescence,

n is the index of refraction of the surrounding solution, κ is dependent upon the geometry of the fluorophore dipole, and J^{λ} is dependent upon the spectral overlap between the donors emission and the acceptors excitation. It is important to note, J^{λ} is not directly found from the spectra overlay, but is defined as $J^{\lambda} = \int f_D(\lambda)\, \varepsilon_A(\lambda)\, \lambda^4\, d\lambda$where $f_D(\lambda)$ is the normalized fluorescence spectrum of donor, $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor and λ is wavelength.

There also exists a geometric relationship between donor and acceptor which affects energy transfer similar to how the orientation of a radio antenna would affect broadcasting. Here kappa, κ ,is introduced to account for the geometry of dipoles such that $\kappa = \left(\hat{d} \cdot \hat{a}\right) - 3\left(\hat{d} \cdot \hat{r}\right)(\hat{r} \cdot \hat{a})$ where $\hat{d}$ and $\hat{a}$ are the directions of the dipole moment of the donor and acceptor respectfully and $\hat{r}$ is the direction from the donor to the acceptor. Kappa can assume values between zero and four. Obtaining accurate kappa values can be done using by measuring the fluorescence anisotropy to minimize errors in the determined distances (252), however this is difficult leading to many experiments assuming a kappa isotropic average of 2/3 based on the high dynamics within the linkers and fluorophores (253). The isotropic average works well when using long linkers to attach the fluorophores to the proteins, but larger errors occur when using short-linkers which restrict mobility. Therefore, longer linkers are advised for FRET experiments, lead to small experimental errors to a maximum of 7% when considering all uncertainties in the measurements (247).

CHAPTER FIVE

PROJECT I: GluN2B DISORDERED C-TERMINAL TAIL AS A POLYMER MODEL

Introduction

Glutamate is the most abundant excitatory neurotransmitter within mammalian central nervous system. Glutamatergic signaling is the primary form of excitatory signaling, making it essential for sensory perception, motor control, cognitive functions, and behavioral traits. Glutamate signaling occurs by binding to and activating receptor molecules which include ion channels and metabotropic receptors. One such family of proteins involved in glutamatergic signaling is the mGluR family. mGluRs are a class of C Glutamate-protein-coupled receptors (GPCRs) which typically form dimers with a large bi-lobed extracellular domain which binds with glutamate (254,255).

The C-terminal domain (CTD) is the largest domain of GluN2B, spanning residues 1259–1482, and is required for GluN2B regulation of N-methyl-d-aspartate receptors (NMDARs) (256). Deletion of the CTD of GluN2B is lethal resulting in the same effect as when the entire until is removed (257). The CTD contains multiple small binding and signaling cites (255,258). Phosphorylation is shown to affect CTD of GluN2B conformational dynamics (259,260). Experimental studies have determined that the CTD of the GluN2B subunit is an intrinsically disordered region (IDR) (258,261). The importance of GluN2B CTD combined with ligand influence on its dynamics makes it a great system to study IDRs including their dynamic behavior and large sampling of

configuration space. Continuing forward, N2B will be used to denote the CTD of GluN2B.

IDRs are thought to be regions containing no secondary structure, moving as a polymer in solution. Low complexity IDRs have been modeled with a simple worm-like chain model (262). However, small changes, such as phosphorylation and the presence of ligands, can alter the dynamic behavior of IDRs while preserving their disordered state, often in unpredictable manners. The intracellular CTD of AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) keeps its disordered behavior when phosphorylated, but increases in rigidity (260). However, phosphorylation within N2B results in a general expansion of the region, resulting in more random coil-like behaviors (259). Prolines are thought to drive disorder; thus, proline depletion should introduce some measure of order into a protein sequence. Often this is done by mutating proline-to-alanine and proline-to-serine. Such a study on N2B shows proline-depleted constructs keep their disordered nature, but underlying dynamics are altered implying global rearrangement of the region's sampled configuration (261 428).

While N2B is classified as an IDR and ensemble measurements imply a globular, disordered state, single molecule Foster Resonance Energy Transfer (smFRET) measurements have shown stochastic conformational transitions on the second timescale (263). Such a result was surprising, implying IDRs sample a large number of conformational spaces rather than random disorder. It seems N2B moves between subsets of conformational states allowing it to keep its fast, disordered dynamics while stochastically altering the available conformational spaces on the second timescale. N2B

seems to ‘hop’ between conformational subsets, terming the phrase hop-diffusion (263). The slow rate of transition between subsets of conformational states implies a large energy barrier separating the subsets with low energy barriers between the local conformational states within the subset (264). Such a structure of the energy barrier is similar to structural glasses which have different energy state accessibility depending upon packing density. Other globular IDRs show similar dynamics and hop-diffusion conformational states, showing N2B is not unique (265).

To better understand the hop-diffusion properties and structural dynamics of IDRs, N2B was taken as a sample IDR. smFRET measurements were taken with eight different labeling distances within N2B. N2B behavior was modeled with worm-like chain and elastic network models, giving insight into the regions inherent chain stiffness.

Methods

Protein Expression/Purification

N2B plasmids were obtained from the Bowen Lab (Stony Brook University, Stony Brook, NY) where GluN2B (residues 1259–1482) from Rattus norvegicus was placed within the expression vector pPROEX HTB (ThermoFisher Scientific, Waltham, MA, USA), which imparts an N-terminal 6-His tag. To enable fluorescent labeling, native cysteines were mutated to serine through classic site-directed mutagenesis as confirmed by DNA sequencing, then cysteine mutations were introduced at the labeling location for each sample. Plasmids were expressed in the Rosetta strain of Escherichia coli (MilliporeSigma, Burlington, MA, USA). Expressed cells were grown in LB media

(10g/L Tryptone, 5g/L Yeast Extract, 10g/L NaCl) at 37*C with 225rpm shaking to an optical density at 600nm (OD) of 0.6, then induced with 1μM of IPTG and allowed to continue growth to an OD of 1.2. The cell pellet was resuspended in lysis buffer (8M Urea, 100mM PBS, 10mM Tris, 1mM BetaMe, pH 8.0) and sonicated to rupture cell wall. The supernatant underwent his-tag purification under denatured conditions using HisPur Ni-NTA Spin Columns (thermoscientific cat#: 88225). Purified proteins were then flash frozen with 10% glycerol until ready for fluorescent labeling.

Labeling

Purified N2B rebound to free Ni-NTA resin and resuspended to a concentration of approximately 10mg/mL for labeling. An equimolar ratio of Atto488 maleimide and Atto647N maleimide was added and left to head-over-tail rock overnight at 4*C within labeling buffer (50mM PBS, 500mM NaCl, 2M Urea, 1mM TCEP, pH 7.5) for random labeling. Unconjugated dye was removed by washing the resin beads with labeling buffer. Labeled protein samples were then eluted from the resin beads by addition of 250mM Imidazole.

Multiparameter Fluorescence Detection (MFD)

Donor fluorophores were directly excited with a 485 nm laser at a power of 75 μW, while acceptor fluorophores were directly excited with a 640 nm laser at a power of 45 μW via pulsed interleaved excitation (PIE) (266). Emission was split into parallel and perpendicular polarization with respect to the excitation beam (267). Labeled samples were diluted to picomolar concentrations within measuring buffer (20mM PBS, 150mM NaCl, pH 7.5, charcoal filtered then 1mM TECEP and 20uM unlabeled protein) such that

~1 molecule per second was observed in the confocal volume. Samples were measured in NUNC chambers (Lab-Tek, Thermo Scientific, Germany) that were pre-treated with a solution of 0.01% Tween 20 (Thermo Scientific) in water for 30 min to minimize protein surface adsorption while measuring. The instrument response function (IRF) was obtained using carbon filtered water. The measuring buffer was used for background subtraction. Calibration experiments and data collection were as previously reported (Ma et al., 2017).

Burst selection was performed using inter-photon arrival time traces to identify single molecules. Burst selection criteria were set such that each burst contained a minimum of 60 photons summed amongst all detection channels, with an inter-photon arrival time cutoff set to the mean minus two standard deviations as measured in the background. The donor fluorescence lifetime and the intensity-based FRET efficiency were calculated for each burst using a maximum-likelihood estimation algorithm (268,269). To ensure both fluorophores were present in each select burst used in MFD histograms, the cutoff values for the difference between observed burst duration in green and red channels under direct excitation of the corresponding fluorophores ($|TGG-TRR|<1ms$) and the observed FRET stoichiometry ($0.3<SPIE<0.7$) was used (268).

Results and Discussion

The FRET network (Figure 5.1), visual representation combination of all labeling sites, has overlapping sections and regions to help gain insight into averaging affects between labeled samples and extract information from overlapping regions. However,

due to the complexity of N2B dynamics, extracting information from overlapping regions was found to be more difficult than expected with common models and approaches not accurately representing the data set. Instead, an overview is provided explaining the differences between DMD simulations and experimentally determined results as well as comparisons to commonly used models for IDRs.

Discrete Molecular Dynamics simulations were performed on N2B. Using Accessible Volume (AV) calculation on the simulation results probing of labeling location combinations can be sampled. AV simulations start by determining the volume available for the labeled fluorophore to sample based on the labeling location, linker properties, and fluorophore geometry. Using the mean distance between the AV of the donor and acceptors, the FRET Efficiencies between fluorophore labeling sites can be determined. A large sampling of various configurations is seen in agreement with the disordered nature of N2B (Figure 5.2). The AV results assisted in building an FRET network (Figure 5.1, Table 5.1) to sample different regions of N2B, leading to the elastic network that is used for modeling.

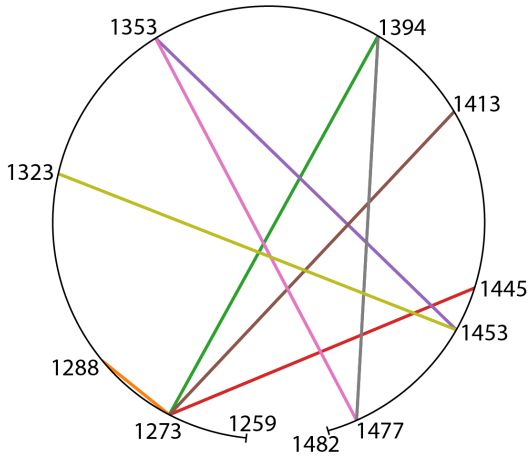


Figure 5.1 **Diagram of FRET Network for N2B samples.** The eight different N2B samples contain different labeling locations as specified by cysteine mutations. In this diagram, each sample is represented by a solid line which connects the labeling locations within that sample. Numbers along the outside of the circle denote the residue number of the labeling site. The break at the bottom of the circle denotes the start of the regions (1259 residue) and end of the region (1482 residue).

Sample	Labeling Sites	Distance Between (residues)
a15	s1273c, s1288c	15
a121	s1273c, c1394	121
a140	s1273c, s1413c	140
a172	s1273c, c1445	172
b130	c1453, s1323c	130
c100	s1353c, c1453	100
c124	s1353c, s1477c	124
d83	c1394, s1477c	83

Table 5.1 **Table of cysteine locations for fluorophore labeling.** For all samples, native cysteines were mutated to serine, unless used for labeling as shown by c####. To introduce labeling sites, serine was mutated to cysteine, as shown by s####c. Where #### represents the full-length residue number.

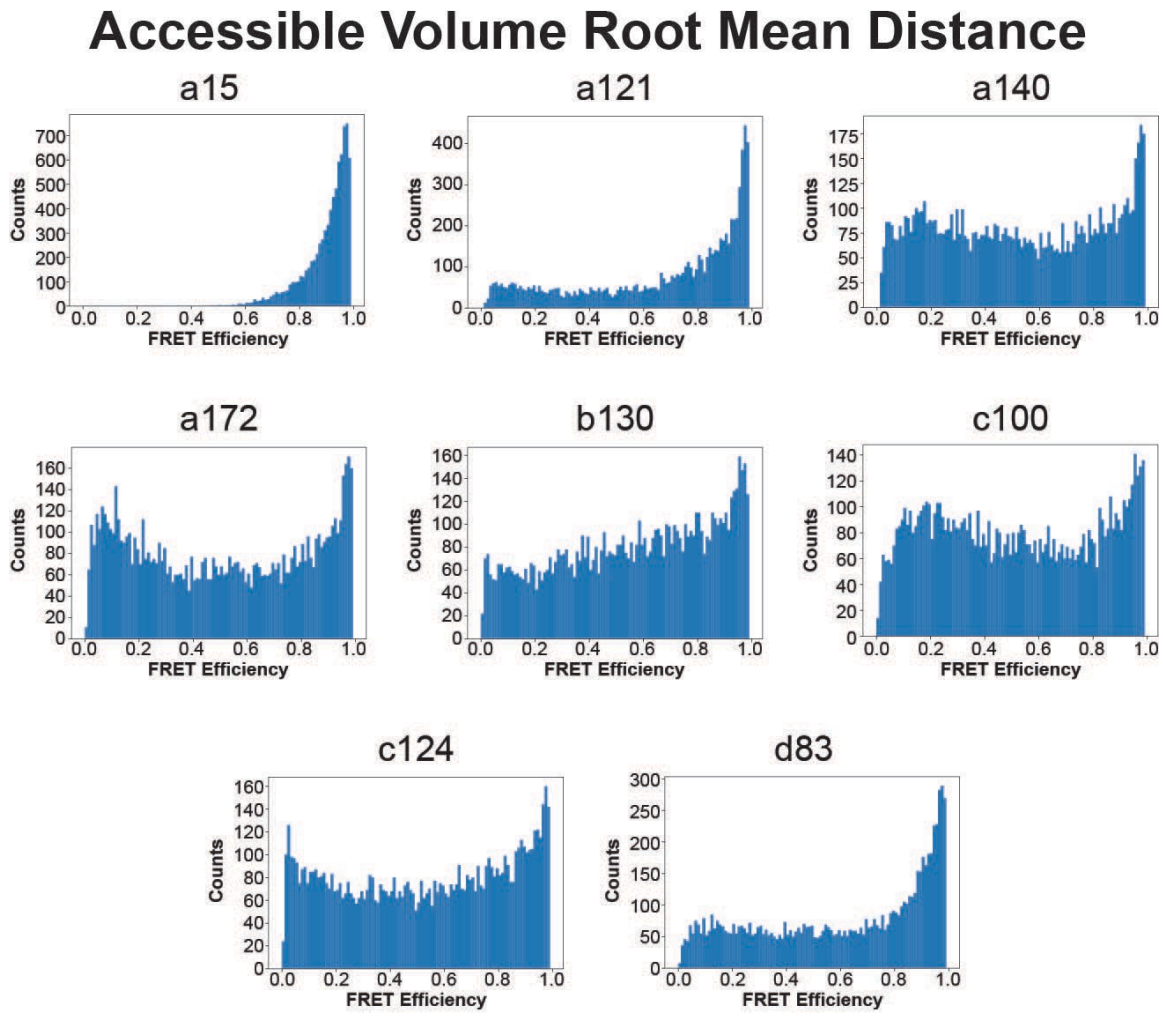


Figure 5.2 **FRET Efficiencies of N2B regions from DMD simulations.** FRET Efficiencies are calculated using the root mean distance between labeling sites obtained from the fluorophore accessible volume (AV) calculations from DMD simulation results.

Multiparameter Fluorescence Detection (MFD) of N2B samples shows obvious populations off the right of the static FRET line (Figure 5.3). Such results indicate dynamic behaviors occurring within the samples. Dynamic population was future explored using time correlated single photon counting (TCSPC) data. Photon arrival times were used to create donor lifetime curves. The mono-exponential (Mono Exp) model was discarded for lifetime fitting, due to not fitting the data as seen in the residuals of the fit. The given lifetime changes seen within the donor seem to fit well with biexponential (Bi Exp) and the worm-like chain (WLC) model (Figure 5.4). However, the WLC model contains on free parameter, the persistence length, while the Bi Exp depends on three free parameters, two lifetime values and the ratio between the population. For this reason, the WLC model gives a more robust fit. Further examination of the χ^2 values for the fits shows the WLC model gives better a slightly closer fit for the data set. Add the WLC modeling to the MFD histogram shows the modeling aligns well with the dynamics seen within the system (Figure 5.3 - blue line). This gives further evidence that the N2B samples show WLC behavior between the labeling sites.

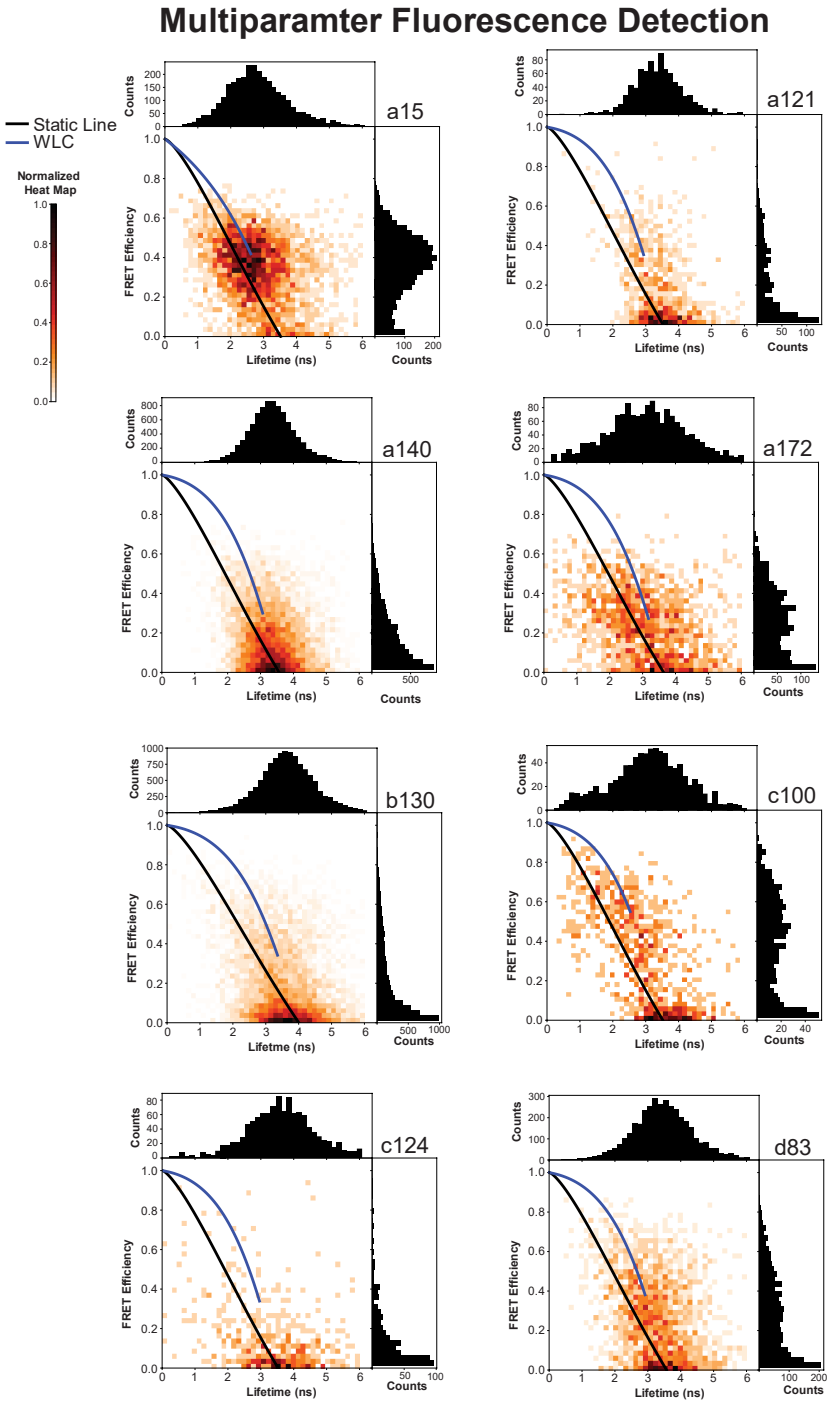


Figure 5.3 **Multiparameter Fluorescence Detection (MFD) data of N2B samples**. The sample identifier is placed at the top, right of each plot. The 1D lifetime histogram is shown along the top, x-axis while the 1D FRET Efficiency histogram is shown along the right, y-axis for each sample. The static FRET line is shown by the solid black line while worm-like chain modeling is shown by the blue line.

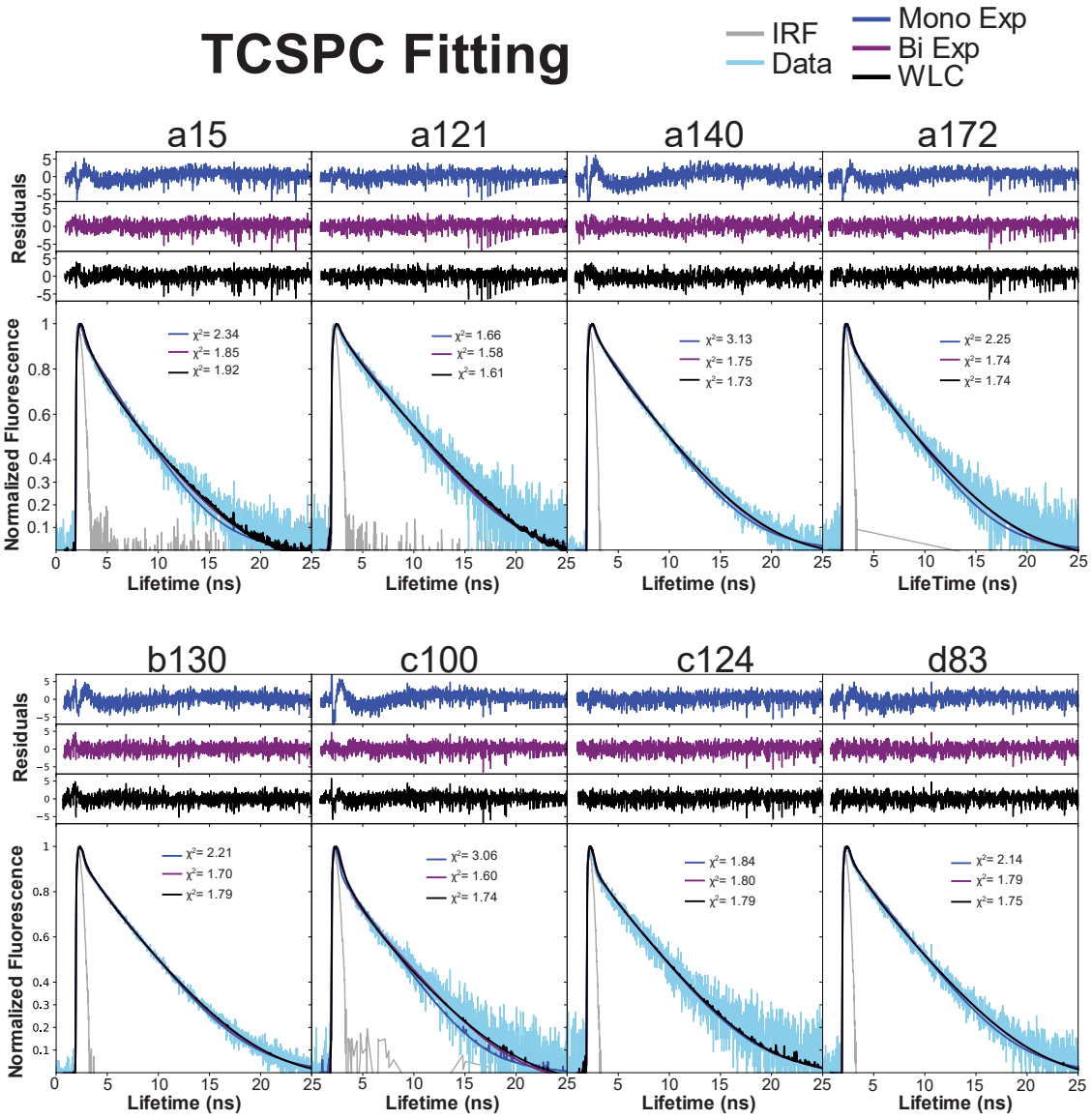


Figure 5.4 **Donor lifetime fits for each N2B sample.** Time Correlated Single Photon Counting (TCSPC) gives lifetime histograms for each sample. Lifetime data has been normalized with background counts removed. IRF and raw data are shown in gray and light blue, respectively. Lifetimes were fits using a mono-exponential (Mono Exp - blue) and bi-exponential (Bi Exp - purple) lifetime model as well as worm-like chain model (WLC - black). Residuals for fitting are shown above the corresponding sample with the same colors as the models. χ^2 for each fit are shown in the top right of each sample graph with the line color indicating the related fit.

Sample	a15	a121	a140	a172	b130	c100	c124	d83
Fit-Nbr	1	1	1	1	1	1	1	1
b_1	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**
bg	**8.146**	**2.43**	**46.278**	**4.589**	**56.346**	**3.144**	**4.405**	**11.184**
dt	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
g	1	1	1	1	1	1	1	1
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**25.315**	**21.456**	**49.766**	30	**23.811**	**11.257**	**12.553**	**39.431**
n0	63824.2	31559.7	404966	36334.9	336296	41854.8	23720.8	104125
r0	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
rep	20	20	20	20	20	20	20	20
rho_1	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**
sc	**0.537**	**0.356**	**0.378**	**0.589**	**0.553**	**0.842**	**0.682**	**0.48**
start	0	0	0	0	0	0	0	0
stop	65.504	65.504	65.504	65.504	65.504	65.504	65.504	65.504
tBg	1	1	1	1	1	1	1	1
tDead	85	85	85	85	85	85	85	85
tL1	**2.853**	**3.252**	**3.326**	**3.089**	**3.612**	**2.714**	**3.632**	**3.445**
tMeas	1	1	1	1	1	1	1	1
ts	**-0.948**	**-0.88**	**-0.989**	**0**	**-0.624**	**-3.683**	**-0.269**	**-0.514**
win-size	17	17	17	17	17	17	17	17
xL1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Chi2r	2.3384	1.6625	3.1339	2.2529	2.2082	3.0609	1.8421	2.1354

Table 5.2 **Fitting Parameters for Mono-Exponential Fit**. Fits were done using ChiSurf. Bolded values were set as free parameters for fitting.

Sample	a15	a121	a140	a172	b130	c100	c124	d83
Fit-Nbr	1	1	1	1	1	1	1	1
b_1	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**	**0.38**
bg	**6.768**	**1.97**	**40.816**	**3.863**	**51.895**	**2.35**	**4.028**	**9.461**
dt	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
g	1	1	1	1	1	1	1	1
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	**12.209**	**9.543**	**36.14**	**138.11**	**18.64**	**13.291**	**11.7**	**35.154**
n0	79077.2	38478	470487	55345.8	371872	93201.4	29127.9	118717
r0	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
rep	20	20	20	20	20	20	20	20
rho_1	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**
sc	**0.356**	**0.235**	**0.267**	**0.298**	**0.46**	**0.203**	**0.495**	**0.366**
start	0	0	0	0	0	0	0	0
stop	65.504	65.504	65.504	65.504	65.504	65.504	65.504	65.504
tBg	1	1	1	1	1	1	1	1
tDead	85	85	85	85	85	85	85	85
tL1	**3.399**	**3.43**	**3.7**	**3.621**	**3.857**	**3.325**	**3.799**	**3.868**
tL2	**1.034**	**0.575**	**1.003**	**0.759**	**1.14**	**0.534**	**0.629**	**1.284**
tMeas	1	1	1	1	1	1	1	1
ts	**-0.547**	**-0.486**	**-0.61**	**0**	**-0.42**	**-1.221**	**0.08**	**-0.161**
win-size	17	17	17	17	17	17	17	17
xL1	**0.563**	**0.738**	**0.697**	**0.531**	**0.796**	**0.315**	**0.755**	**0.694**
xL2	**0.437**	**0.262**	**0.303**	**0.469**	**0.204**	**0.685**	**0.245**	**0.306**
Chi2r	1.8513	1.5832	1.751	1.7372	1.6961	1.595	1.7993	1.794

Table 5.3 **Fitting Parameters for Bi-Exponential Fit.** Fits were done using ChiSurf. Bolded values were set as free parameters for fitting.

Sample	a15	a121	a140	a172	b130	c100	c124	d83
Fit-Nbr	1	1	1	1	1	1	1	1
R0	52	52	52	52	52	52	52	52
R0 [A]	51	51	51	51	51	51	51	51
bg	**1.839**	**-0.277**	**39.016**	**3.832**	**48.910**	**2.278**	**4.002**	**9.318**
donly	**0.154**	**0.328**	**0.325**	**0.157**	**0.411**	**0.055**	**0.827**	**0.434**
dt	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
g	1	1	1	1	1	1	1	1
kappa2	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667
l1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
l2	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
lb	3.055	3.060	34.132	20.100	21.080	13.851	11.599	34.099
length	**94**	**475.6**	**544**	**659.2**	**508**	**400**	**486.4**	**338.8**
link_ width	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
n0	76400	35074	453720	44085	396065	63588	27292	190376
persis-tence	**24.53**	**6.08**	**5.66**	**4.51**	**5.33**	**3.38**	**3.87**	**6.72**
r0	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
rep	20	20	20	20	20	20	20	20
sc	**0.116**	**0.136**	**0.148**	**0.208**	**0.262**	**0.092**	**0.515**	**0.212**
start	0	0	0	0	0	0	0	0
stop	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52
tBg	1	1	1	1	1	1	1	1
tDead	85	85	85	85	85	85	85	85
tL1	**4.264**	**3.937**	**3.952**	**3.992**	**4.104**	**4.006**	**3.839**	**3.993**
tL2	**1.671**	**1.382**	**1.243**	**1.506**	**1.468**	**1.568**	**1.130**	**1.672**
tMeas	1	1	1	1	1	1	1	1
tau0	3.52	3.525	3.565	3.64	4.005	3.52	3.525	3.57
ts	**0.034**	**-0.469**	**-0.482**	**-0.581**	**0.00017**	**-1.715**	**0.025**	**0.092**
win-size	17	17	17	17	17	17	17	17
xL1	**0.783**	**0.829**	**0.857**	**0.876**	**0.889**	**0.840**	**0.827**	**0.818**
xL2	**0.217**	**0.171**	**0.143**	**0.124**	**0.111**	**0.160**	**0.173**	**0.182**
Chi2r	1.9173	1.6068	1.7253	1.7444	1.7897	1.741	1.7925	1.7454

Table 5.4 **Fitting Parameters for Worm-like Chain model.** Fits were done using ChiSurf. Bolded values were set as free parameters for fitting.

Sample	Residues between Labeling Sites	Region Length (A)	Persistence Length (A)
a15	15	94	24.53
a121	121	475.6	6.08
a140	140	544	5.66
a172	172	659.2	4.51
b130	130	508	5.33
c100	100	400	3.38
c124	124	486.4	3.84
d83	83	338.8	6.72

Table 5.**5 Parameters for elastic network.** Region length is dependent upon the number of residues between labeling locations. Persistence length is taken from worm-like chain fitting.

Segments	Start AA	End AA
Start	1259	1273
A	1273	1288
B	1288	1323
C	1323	1353
D	1353	1394
E	1394	1413
F	1413	1445
G	1445	1453
H	1453	1477
End	1477	1482

Table 5.6 **Segment regions for elastic model fitting.** Segmented regions are created according to labeling locations and overlapping regions.

However, when applying the elastic next work model, results remain inconsistent and not biologically relevant. Thus, while the labeled samples fit a WLC model, stiffness of overlapping regions cannot be extracted from the current data set using the spring in series model (eqn. (5.1)). For eqn. (5.1), all text variables represent the stiffness of that region. The lack of appropriate results for the spring in series model could be due to an averaging effect that occurs during measurements that alters results between overlapping regions.

(5.1)

$$\begin{bmatrix} 1/a15 \\ 1/a121 \\ 1/a140 \\ 1/a172 \\ 1/b130 \\ 1/c100 \\ 1/c124 \\ 1/d83 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1/A \\ 1/B \\ 1/C \\ 1/D \\ 1/E \\ 1/F \\ 1/G \\ 1/H \end{bmatrix}$$

Since the elastic network model does not appropriately fit the experimental results, further data analysis was done to determine what polymer models might more accurately fit the data set. Polymers with a uniform stiffness will show persistence length that scales linearly with length of the polymer (270,271). Exploring this concept (Figure 5.5 **N2B polymer modeling comparison.** A) Semiflexible chain modeling fitting for N2B. The outlier from samples a15 (red) is excluded from fitting. B) Extended polymer modeling as depending upon FRET efficiency. A fully extended polymer (black) fits sample a15 (filled with red) but excludes the other samples. Experimental (orange) and

AV simulations (blue) show different trends with, neither of which fits the model.Figure 5.5A), the N2B samples do not show a well-structured linear fit. Sample a15, with only a 15 amino acid length, is an obvious outlier for the data set (shown in red). Due to its short nature, the sample is too short to accurately fit such polymer models. When the outlier is removed from the data set, the remaining samples give a stiffness of 0.118 with length being equivalent to the number of amino acids multiplied by the average length of an amino acid, 3.6 Å.

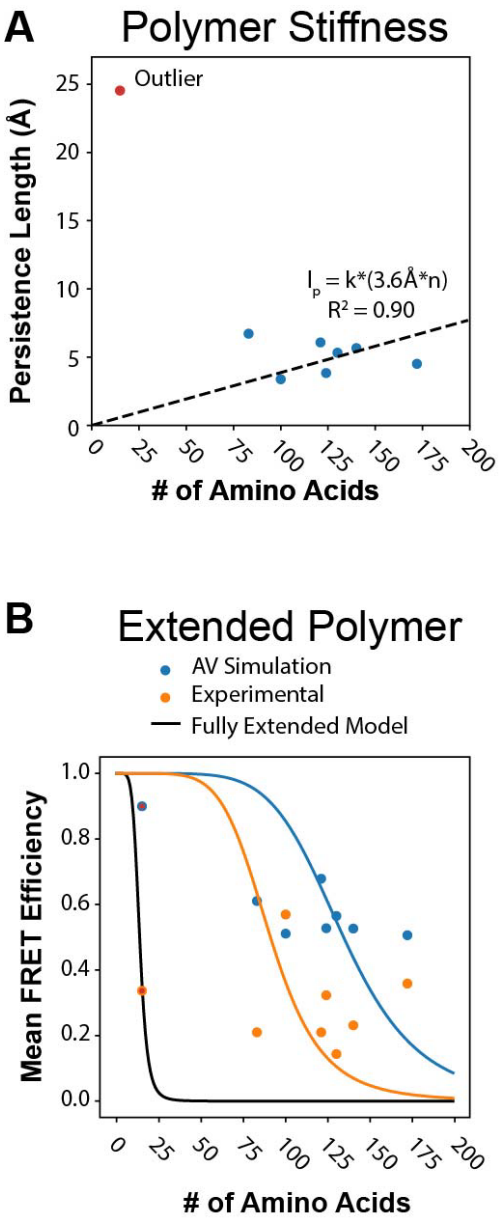


Figure 5.5 **N2B polymer modeling comparison.** A) Semiflexible chain modeling fitting for N2B. The outlier from samples a15 (red) is excluded from fitting. B) Extended polymer modeling as depending upon FRET efficiency. A fully extended polymer (black) fits sample a15 (filled with red) but excludes the other samples. Experimental (orange) and AV simulations (blue) show different trends with, neither of which fits the model.

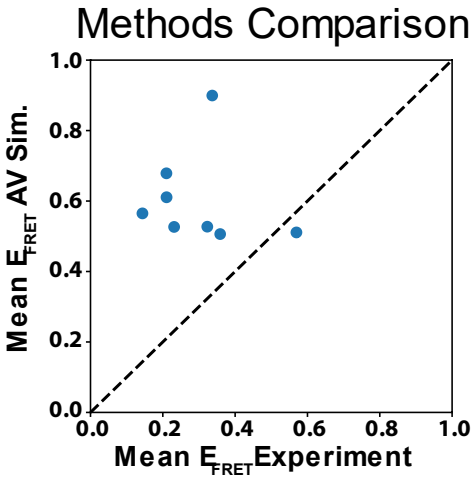


Figure 5.6 **Comparison of experimental and AV simulated mean FRET efficiency.** Equal values are shown with the dashed, black line.

Further, N2B does not fit a fully extended polymer model (Figure 5.5B). When using 3.6 Å for a fully extended protein region (black), only the shortest sample, a15, with only 15 amino acids fits the curve. Fitting the model with the length of an amino acid being a free parameter gives 0.365 Å and 0.538 Å as a single amino acid length for simulated and experimental data, respectively. This indicates the experimental results are showing a a larger measure of expansion than seen on the simulation. This result shows that N2B collapsed, agreeing with N2B being a globular IDR, as previously mentioned.

Simulation data show higher FRET efficiency, thus closer distances, than seen in the experimental results (Figure 5.6). Looking into the discrete molecular dynamics (DMD) for N2B, briefly sampled many configurations, as expected for an IDR. However, there are very apparent secondary structures arising, unexpected for an IDR (Figure 5.7). The conformational states are very short lived, rapidly sampled giving N2B the fast dynamics as expected from disorder. The presence of second order structural configurations would explain why homogenous polymer models do not fit the system. Further evidence is shown by PONDR predictions where the majority of samples are on the borderline of order and disorder with values between 0.4 and 0.6 (Figure 5.8). As disorder increases, the persistence length is expected to decrease, but no clear trend is seen within the data set.

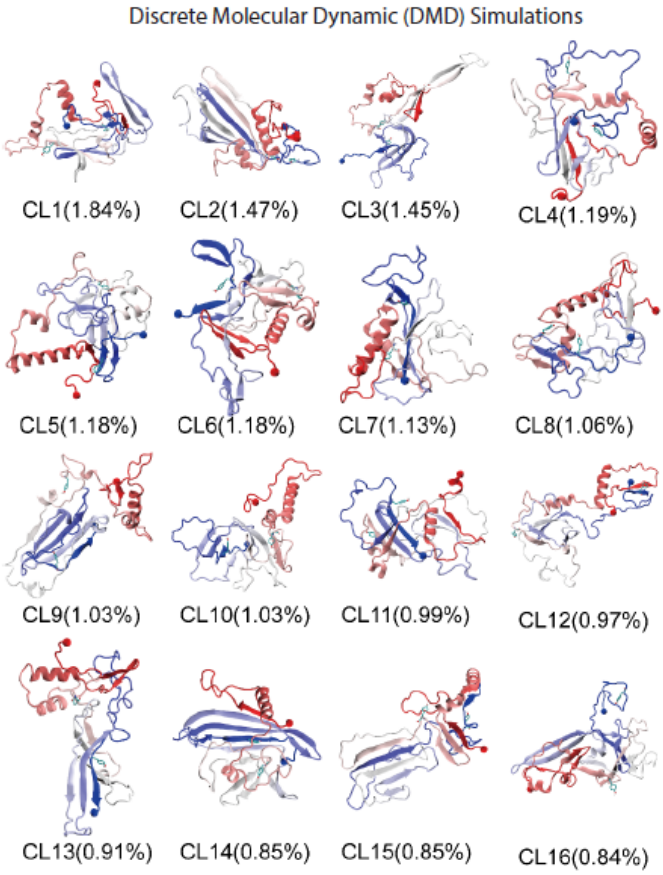


Figure 5.7 **Sampled configurations of N2B from DMD simulation.** These are only from the 12 largest clusters.

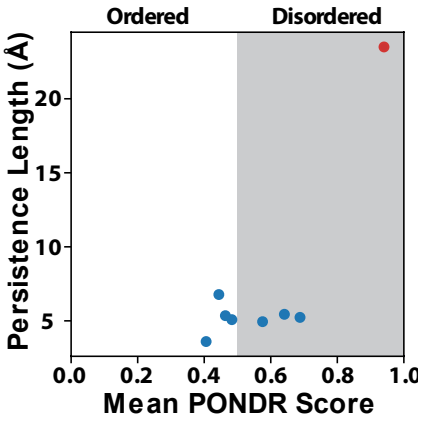


Figure 5.8 **Persistence length compared to PONDR prediction.** Outlier (red) is sample a15, only 15 amino acids in length.

Conclusion

N2B is a disordered region with high dynamics. Data analysis on independent samples shows worm-like chain (WLC) modeling gives the most robust fitting for lifetime data and aligning nicely with the FRET populations along the 2D histogram, despite the elastic network model not working for the entire region. In contrast to traditional IDR models, N2B shows secondary structures during DMD simulations, but these structures are sampled rapidly with no appearance of a stable state. The presence of secondary structure could explain the discrepancy between polymer models, such as the elastic network, in terms of persistence length and labeling site distances dependencies. The addition of secondary structures would modify the persistence length due to the dependency upon secondary structures and interactions with persistence length. This results in the region between labeling sites not having a constant persistence length as expected for the elastic network model, but rather it is a combination of various contributions of secondary structure like states. However, worm-like chain modeling may fit individual samples well due to an averaging effect occurring along the secondary structures due to rapid movements which leads to features of the secondary structures becoming lost. These same averaging effects could explain the lack of relationship between disorder and persistence length between labeling sites, since more ordered regions would be predicted to have a higher persistence length due to the nature of alpha helix and beta sheet formations. Further experiments need to be completed to give further insight into such systems.

Overall, N2B is not a typical IDR. The basic worm-like chain modeling for N2B, which assumes N2B is a non-interacting polymer, does not appropriately fit the experimental data when all samples and overlapping regions are considered. The rapid sampling of secondary structures can be explained, as seen in the simulations, as numerous weak intrachain interactions. It is reasonable and rational that such a 'polymer' would have interactions with itself, as electrostatic forces, and the surrounding environment, as hydrophobicity affects. These weak interactions lead to a rough energy landscape, similar to that of an IDR (264). While the current understanding in the field is to treat IDPs as polymers, specifically worm-like chains, there is little being done to currently expand the polymer models to include long-distance electrostatic forces or secondary structures. The strength and overall contribution of these weak interactions is yet to be determined. However, these results show the need for a different manner of IDR modeling, one that considers rapid structure sampling, rather than only random coil interactions.

Future Work

Future experiments include varying the buffer conditions for N2B to try and determine the contributions from weak intramolecular interactions. Such experiments include increasing the salt concentration negating any self-interactions due to electric forces, adding a denaturant to remove any dynamic restrictions due to secondary structures, and adding a known ligands and binding partners for signaling to understand how binding for signaling play a role in dynamics and structure. Together, these results

allow a glimpse into understanding the dynamics and structure of intrinsically disordered proteins.

Recent work has shown N2B can participate in phase separation (272). This is not surprising, since IDRs are predicted to drive phase separation. However, not all IDRs can phase separate, such as GluN2A CTD. Understanding the unique properties of N2B and comparing it to other IDRs that are known to phase separate may help identify other IDRs that contain unstable secondary structures. Further, biomolecular condensates are highly dependent upon environmental aspects. By understanding the influence of environment on a single IDR and how intramolecular forces are affected by environment, it can help understand how intermolecular forces can play a role biomolecular condensate formation.

CHAPTER SIX

PROJECT II: GLUTAMINE DEPENDENCY ON BIOMOLECULAR CONDENSTATE STABLITY

Introduction

To function, the cell must have a system of organization and separation. Historically, this has been thought to occur within membraned organelles with specific cellular functions. Currently, it is known that such organization can occur without use of a membrane through the formation of biomolecular condensates (23,273). Biomolecular condensates offer several benefits for the cell including environmental response for signaling and regulation (29,48,49,274) as well as storage and organization (25,28,38,51). Biomolecular condensates have liquid like properties as seen from FRAP experiments (29,39,275-277) implying proteins forming biomolecular condensates remain highly dynamic within the condensate and participate in exchange with the surrounding cell cytoplasm (19, Wei, 2017 #82,23). Current evidence suggests biomolecular condensates form through a similar driving force as liquid-liquid phase separation where the protein are modeled as a polymer through Flory-Huggins theory.

Many proteins that drive liquid-liquid phase separation contain low complexity (LC) intrinsically disordered regions (IDR) (36,41-44). Evidence shows that proteins containing IDR linkers between structured domains can form biomolecular condensates while leaving the structure domain largely intact (44-46), implying proteins with LC IDRs are possible candidates for BC formation. This is shown to be due to the multiple interaction locations within a single LC-IDR allowing for many dynamic, weak

interactions to occur at once, building a network that drives the proteins together into a liquid phase. This makes multivalence a major factor for biomolecular condensate formation and leading to the most promising modeling technique for biomolecular condensate formation prediction, the sticker-and-spacer model (44,47,53,113).

Biomolecular condensates are highly important within the cellular nucleus involving proteins, DNA, and RNA (27,51,278,279) with the reoccurrence of Glutamine Rich (Qrich) regions, which are LC IDRs, often found within transcription factors, which are essential for cellular function. Typically, Qrich and long repeating glutamine regions (PolyQ) are associated with the formation of harmful aggregates through Beta-sheet interactions that lead to various diseases such as Huntington's and neurological disorders. Recent studies have shown beta sheets may play a larger role than expected in biomolecular condensate formation and stability (108). This has led us to investigate Glutamine's role and contribution to biomolecular condensate stability.

The Flory-Huggins theory for mixing polymers is the main model used to predict phase separation for proteins (36,176). Proteins are modeled as polymers allowing for the Flory-Huggins interaction parameter for phase separation to be determined using the Flory-Huggins theory (180,182). The Flory-Huggins interaction parameter is a measure of the free energy of mixing for mixed solutions and accounts for enthalpic and entropic changes in solubility. Such modeling gives a good prediction for if a protein system will phase separate (176,178,182). The critical temperature gives a measure of biomolecular condensate stability as higher temperatures are required for the entropic forces to overcome the enthalpy. The Flory-Huggins theory only requires the volume fraction of

protein in the dense and dilute phases to build a phase diagram for phase separation, making it a simple model to use to predict phase separation based off several data points, rather than the multiple data points required to build an entire phase diagram curve.

In this study, we use previously established coarse-grained models, simulated with LAMMPS (205) to predict biomolecular condensate formation at various temperatures for varying glutamine percentages within a 100 amino acid chain. Five 100 amino acid regions are selected from Human Transcription factors containing 81%, 77%, 65%, 54% and 40% glutamine. The wild type (WT) sequence along with three randomly scrambled sequences were coarse grained via amino acid, then simulated at varying temperatures to obtain a protein concentration within the dilute and dense phase. These protein concentrations are then used for Flory-Huggins modeling to predict the critical temperature and create a phase diagram for the sample. The dependency of glutamine percentage on biomolecular condensation formation and critical temperature are found to be indirectly related with higher glutamine percentages leading to lower critical temperature. This may be due to the manner of the sticker-spacer model where the addition of glutamine, typically modeled as a spacer, displaces other amino acids which act as stronger stickers. However, the aggregation of Qrich regions is not displayed within the simulations suggesting that the lower stability of a dense liquid phase may offer a lower liquid to solid energy barrier for higher Q regions, but that remains to be seen.

Methods

Sample Selection

UniProt was used to obtain a list of known human transcription factors. Since Qrich regions are not well defined, an in-house python script was used to determine the highest glutamine (Q) percentages for a 100 amino acid region of the proteins using a sliding 100 amino window for determination. Five different amino acid regions were selected based on varying glutamine percentages (Q%) within the sequence. Selection was done in a manner to encourage an even spread between the maximum Q% being 81%, and a low cut off of 40% (Table 6.1). The original amino acid sequence, denoted by O, where scrambled randomly using an in-house python script. The first scramble, denoted by A, was scrambled again to obtain sequence B, which was scrambled again to obtain sequence C (Table 6.2). Samples are named with the letter of the scramble, followed by a two-digit number expressing the glutamine percentage for the region.

Percent Q	Protein Code	Protein Name	Amino Acid Region
40%	Q9Y6Q9	Nuclear receptor coactivator 3	1208 – 1307
54%	G3V1P5	Mediator of RNA polymerase II transcription subunit 15	64 – 163
65%	O14686	Histone-lysine N-methyltransferase 2D	3869 - 3968
77%	Q93074	Mediator of RNA polymerase II transcription subunit 12	2049 – 2148
81%	D3VVP5	Ubiquitinyl hydrolase 1	280 – 379

Table 6.1 **Protein regions selected for study.** Each region is 100 amino acids in length with varying glutamine (Q) percentage.

Name	Amino Acid Sequence
O40	PMMQPQVSSQQGFLNAQMVAQRSRELLSHHFRQQRVAMMMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQTQAFSPPPNVTASPSMDGLLAGPTMPQAPPQ
A40	PQMQQSMRAQMQRQHAQMFQNFGTQVGQQPQSQNVAQTQQMPSLLQRGMQSQPQDQQESQPQAFQQMQQPLPQSQLQAQQSQMQQPLQAQPPHRTAQQV
B40	QNMNQQTQFQQQQHAALQGSPMMTPMLAQPQQMQMHSGQQLVQFAPAQSQPQQRQPTGSQFPVSVSPQMSQQQRQRQLVQQQQRAQQQPQQQMPADLEQQ
C40	QAREPPQLAQMTQQQQRPQLQALHSPVQTMNVAQQQPPQHQSQQGLQPMSQQNQGQPQPASMVQMQVMQQQLDQQQFSQMAAQQQQQPQRMSFTQSRGFQ
O54	LQQQQQQQHLIKLHHQNQQQIQQQQQQLQRIAQLQLQQQQQQQQQQQQQQQALQAQPPIQQPPMQQPQPPPSQALPQQLQQMHHTQHHQPPPQPQQPP
A54	HPHHHQQQQPPAQQLPQIPHPQLLILMLPQQPQQQQATQQQMHQKAQQQQLQQQQLQQPQPQLPQIHQQQQQQQPQLQQQQPAPQPQQRSQQNQIQQQQ
B54	HQQLQQPAMQHKPPPQHQAQQQPQQMQNQPPLIQLQLPAQPQIQQQQQQQLQQLLQQQRSQHQQILQQHAPQQQPQQHIQQTQHPPQQQQPLQQQQQQPQ
C54	HQQQQQPQPHHPPQHHQPQIAQMQQIQQLLNAIQQQQQQQQLQPQQQQQSLQLQQLQQRQTQLQPQLHHAAQPQQPQKQMLQIPPQQQPPQPQPQQQQQQ
O65	SMAGLSHLQQSLMSHSGQPKLSAQPMGSLQQLQQQQQLQQQQQLQQQQQQQLQQQQQLQQQQLQQQQQQQQLQQQQQQQLQQQQQQLQQQQQQQQQQFQQ
A65	QQQQMLKQQALQLQQQQQQQQQQSQLQQQLQQQMQGQQQPQQQLLQQQQLQPQQGQQLQQQSSQQQQLQLQSQQQQQGLQSLQQHLQQQQAHQSMQQSFQ
B65	QQLQQSLQQGQQQQQQQHQQQFLQQQQQSQQQQQQQSLGQQQQQQQQAQLPLQQQQQQQQSQMQQLQKSLQQQLSQLHQQQMMLPLQQQQQGLQQLQSQA
C65	QQGPLQQMQSLQQQQQQLQLGQQQGQQAQQSQQQQQQQHQQQQQQQLQSQQLQQQQLQQQQLQQLMQALQQQQKPQQQQQQHQQSSQLQLQLMLQSSFQQ
O77	PEQQQQQQQQQQQQQQQQQQQQQQQQQQQQYHIRQQQQQQILRQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQQQQAAPPQPQPQSQPQFQRQGLQQTQQQ
A77	QQLQQQQPQQPQHQRPQQQQAQQQQPQQQQQQQQSQQQQQQQQFEHQQAQQQQQQPQQLQQQQQQQQQTQRQQQQQQQQQQQQQQIQQQQQQQGYQPQRIQ
B77	QQRQQQQQQQQQQQHQQQQQQPIQQQTQQQAGQQQQRQQAQQQQQLFQQQQHQQQPQQQQQQPQQQQQQQISQQQQQLQQQRQQQYQQQQPPEQQQQPQQ
C77	QQQQQQQGQQQLQQQQAPPQQQQQQRQQYRSQQQQQQHQQRQQQQQQQQQQQQQQQQQQHIQQLQTQAQQQQQIQQQQFPEQQQQQQQQQQQQQQPQQPP
O81	ELRKRREAYFEKQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQRDLSGQSS
A81	QQQQQQRDQQLQQQQQQQQQQQQQQQSAQQQQRQQQQQSEQFQQQGQQQQQKQQQQQQQEQEQQKQRQQQQQQQQQQQSQQQQQQQQQQQQQLQQYQQQQQQR
B81	QQQQRQQQSQQQSQGQQQQQQLQQQQQKQQQQQASQFQQQQQDYQQQQQQQQEQLRQQQQQQQQQQQQQQQQQQQQQQQQQQEQQQKQQQQQQRQQQQREQQ
C81	QQQQQQQQQQQKQQQQQQRQQLQQQQQQQQQQQQQQQLQEQQQQQGQQQQQQQQQQQQQRQQRQQRQQEQQQQFQEQDQKQQQQSQQQQQSQSQQQQQYAQ

Table 6.2 **List of sequences used for study.** The samples are names from their scramble and glutamine percentage. O indicates the original sequence, while A, B, and C are scrambles from the original.

Simulations

Due to the disordered nature of Qrich regions, only sequence information was used to create a coarse grain input file giving a starting location of amino acids for Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) (201) scripts. Using an in-house python script, one hundred identical chains of the selected 100 amino acid region were randomly placed, fully extended along z-axis, with a 200nm-by-200nm-by-200nm box and the parameter input file including the amino acid intermolecular interaction parameters and temperatures for the simulation was created. For the initial input file, each amino acid was modeled as a bead with a specific mass, charge, and interaction parameters as previously described (Table 6.3) (184).

LAMMPS force fields modeled electrostatic interactions using a Coulombic term with Debye-Huckel electrostatic screening (280) to account for implicit solution conditions such as salt levels. Amino acid interactions were modeled with hydrophobicity (λ) and sizing (σ) considerations using Ashbaugh-Hate functional form which has been shown to work well for disordered proteins (eqn. (6. 1)).

(6. 1)

$$\Phi(r) = \begin{cases} \Phi_{LJ} + (1-\lambda)\epsilon, & \text{if } r \leq 2^{1/6}\sigma \\ \lambda\Phi_{LJ}, & \text{otherwise} \end{cases}$$

$$\Phi_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Amino Acid Type	Mass (amu)	Charge	σ (Å)	λ
ALA	71.08	0	5.04	0.730
ARG	156.20	1	6.56	0.000
ASN	114.10	0	5.68	0.432
ASP	115.10	-1	5.58	0.378
CYS	103.10	0	5.48	0.595
GLN	128.10	0	6.02	0.514
GLU	129.10	-1	5.92	0.459
GLY	57.05	0	4.50	0.649
HIS	139.10	0.5	6.08	0.514
ILE	113.20	0	6.18	0.973
LEU	113.20	0	6.18	0.973
LYS	128.20	1	6.36	0.514
MET	131.20	0	6.18	0.838
PHE	147.20	0	6.36	1.000
PRO	97.12	0	5.56	1.000
SER	87.08	0	5.18	0.595
THR	101.10	0	5.62	0.676
TRP	186.20	0	6.78	0.946
TYR	163.20	0	6.46	0.865
VAL	99.07	0	5.86	0.892

Table 6.3 **Interaction parameters for LAMMPS simulation.** Each amino acid has its own specified radius (σ), mass, charge, and hydrophobicity (λ) parameter used to calculate the force field. Values are the same as previously used (184).

Res	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	-0.042	-0.123	0.054	0.072	-0.239	-0.004	-0.013	-0.218	0.091	-0.249	-0.157	0.041	0.023	0.036	0.041	0.025	-0.005	-0.167	-0.146	-0.103
C	—	-0.299	-0.013	0.	-0.333	-0.084	-0.125	-0.305	0.03	-0.336	-0.256	-0.03	-0.075	-0.055	-0.028	-0.056	-0.079	-0.254	-0.253	-0.178
D	—	—	0.1	0.118	-0.114	0.064	-0.005	-0.085	0.056	-0.107	-0.028	0.056	0.089	0.076	-0.002	0.06	0.044	-0.02	-0.054	-0.046
E	—	—	—	0.128	-0.122	0.099	0.011	-0.094	0.044	-0.124	-0.058	0.072	0.095	0.08	0.	0.074	0.05	-0.038	-0.068	-0.049
F	—	—	—	—	-0.47	-0.175	-0.236	-0.431	-0.103	-0.472	-0.404	-0.14	-0.187	-0.173	-0.161	-0.165	-0.19	-0.379	-0.367	-0.32
G	—	—	—	—	—	0.003	0.011	-0.142	0.106	-0.178	-0.106	0.05	0.038	0.058	0.052	0.042	0.018	-0.105	-0.108	-0.07
H	—	—	—	—	—	—	-0.074	-0.176	0.087	-0.214	-0.161	0.018	0.002	0.027	0.01	0.015	-0.014	-0.124	-0.161	-0.118
I	—	—	—	—	—	—	—	-0.403	-0.07	-0.45	-0.354	-0.091	-0.14	-0.132	-0.128	-0.118	-0.166	-0.356	-0.331	-0.281
K	—	—	—	—	—	—	—	—	0.203	-0.104	-0.02	0.1	0.123	0.092	0.158	0.115	0.091	-0.021	-0.04	-0.031
L	—	—	—	—	—	—	—	—	—	-0.481	-0.39	-0.139	-0.182	-0.167	-0.166	-0.156	-0.195	-0.397	-0.365	-0.321
M	—	—	—	—	—	—	—	—	—	—	-0.301	-0.064	-0.111	-0.097	-0.08	-0.072	-0.117	-0.288	-0.309	-0.249
N	—	—	—	—	—	—	—	—	—	—	—	0.056	0.07	0.053	0.059	0.065	0.037	-0.053	-0.075	-0.046
P	—	—	—	—	—	—	—	—	—	—	—	—	0.049	0.051	0.054	0.066	0.035	-0.099	-0.138	-0.087
Q	—	—	—	—	—	—	—	—	—	—	—	—	—	0.069	0.044	0.074	0.035	-0.075	-0.079	-0.066
R	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.068	0.061	0.035	-0.075	-0.107	-0.084
S	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.057	0.029	-0.074	-0.068	-0.048
T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.014	-0.112	-0.09	-0.07
V	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	-0.306	-0.274	-0.222
W	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	-0.263	-0.225
Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	-0.179

Table 6.4 **Interaction parameters between amino acids.** Taken from (184).

LAMMPS simulations where run with 10 fs steps with implicated solvent considered by a dielectric constant of 80.0 representing water. Velocity assignment of amino acids is completed according to a Boltzmann distribution for the assigned temperature. All boundary conditions remain period over the course of the simulations. The LAMMPS steps are as follows and visualize in (Figure 6.1):

1) Equilibration
    a. Since the initial input file has the amino acids as a fully extended polymer, the system is allowed to relax into a 'native' state. This is done over 50,000 time-steps under constant Number, Pressure, and Temperature (NVT) conditions with temperature of 150K.
2) Compression
    a. After the amino acid chains have relaxed, the volume of the compressed over 25,000 time-steps, forcing the amino acid chains into a dense, compressed droplet at 150K.
3) Expansion
    a. Still at 150K, the z-axis of the box is expanded to 100nm over 50,000 time-steps to allow empty space in which the chain may expand into. Remaining at a low temperature shows slight droplet expansion, but all samples remained in a high density state with no chains drifting out of the droplet.

4) Temperature Increase

a. At constant number and pressure conditions, the input temperature is linearly increased from 150K to the temperature of interest over 100,000 time-steps. This is done by assigning new velocity magnitudes following a Boltzmann distribution every couple time steps. This ensures a graduate increase of temperature for the molecules and eliminates non-pressure errors caused by a sudden increase in temperature. During this stage, amino acid chains show increased dynamics as expected the associated velocity increase. The droplet expands and chain may disassociate with the droplet, creating a low-density phase.

5) Equilibration

a. The simulation continues running at the desired temperature under NVT conditions for another 2,500,000 time-steps, ensuring equilibrium is reached for the system.

The simulations result in a thermodynamic dump file containing the potential energy, kinetic energy, temperature, box dimension in x, y, and z, as well as the pressure tensor for every 1,000 time-steps. This file is checked for stability, ensuring equilibrium of the system is reached.

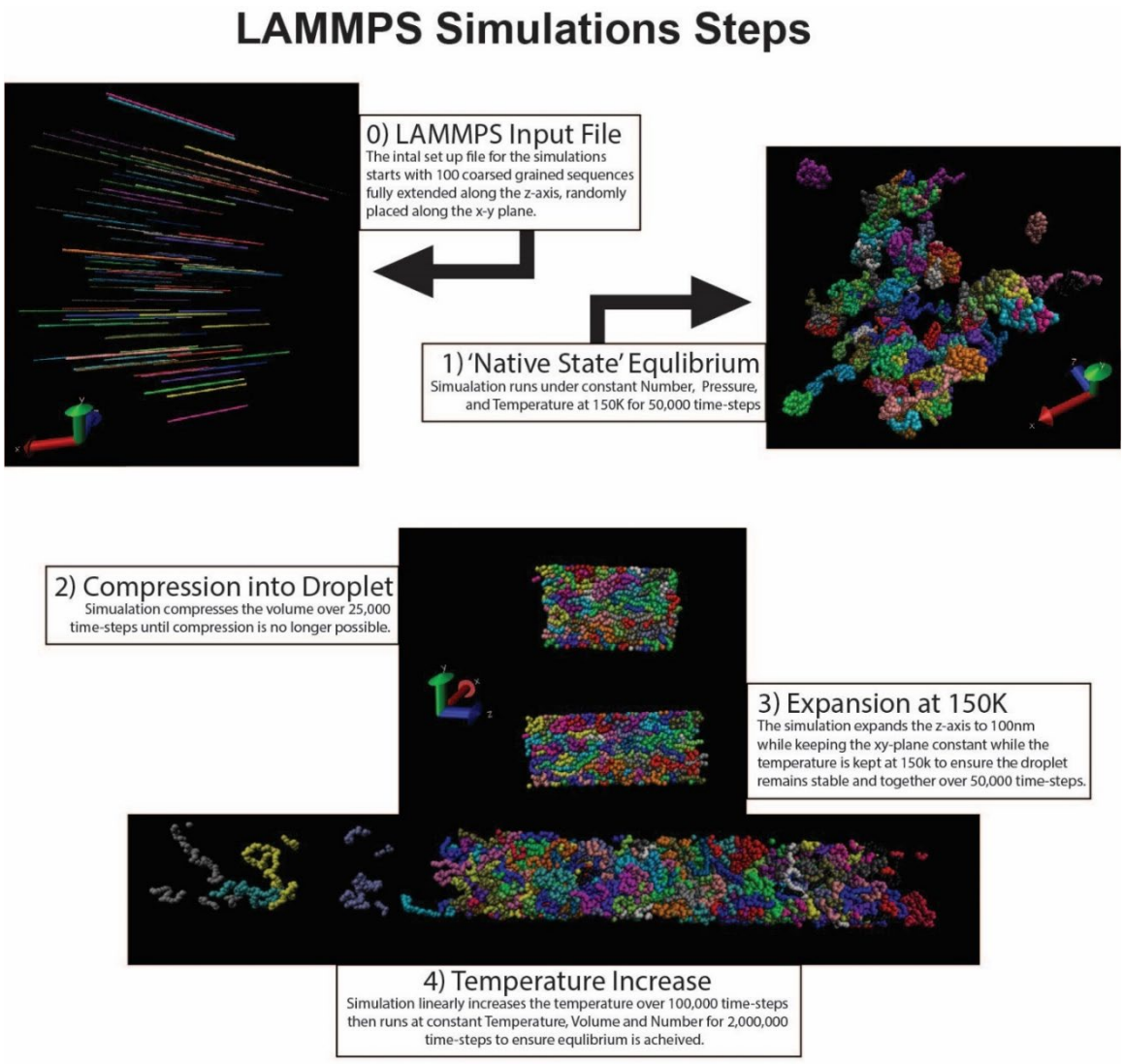


Figure 6.1 **Flow chart for LAMMPS simulation steps.** Visualization of simulation steps are shown using VMD where each amino acid chain is colored differently. This example is from sample O65 at a temperature of 350K.

Obtaining Densities

Every 1,000 time-steps all amino acid locations are recorded into an xtc file. The xtc file is loaded into Visual Molecular Dynamics (VMD) (281), which is used to visual the simulation, ensure all stages have completed as expected, and determine the density of chains along the z-axis using the add on plug in 'Profile Density Tool' (282). Profile Density Tool determines the density by slicing the xy-plane into 1nm even slabs along the z-axis and counting the amino acids within that region. The density of the last 2,000 frames (2,000,000 time-steps) is then averaged for each slab giving a density profile of amino acid distribution along the z-axis. High- and low-density phase regions were manually selected and averaged to give the density for each phase.

Data Fitting

After the set of high- and low-densities at varying temperatures are obtained for each amino acid sequence, the densities are input into a python script for Flory-Huggins Phase Diagram fitting supplied by Brady (37). Brady's script starts by converting the density within a phase, ρ, into the volume fraction of protein within the high- and low-density phases. This is done by introducing a free parameter, γ, to represent the size of the protein such that the volume fraction, Φ, is given as

(5.2)

$$\phi = \frac{\rho}{\gamma}$$

The interactions parameter, χ, is a linear function with respect to inverse temperature (Chapter 3.3) such that

(5.3)

$$X = A + \frac{B}{T}$$

Where A and B are sequence dependent parameters. Within coarse-grained simulations, the entropic contributions at absolute zero will be negligible, thus finding the best fitting for χ is when the y-intercept, A, is zero. Following the equilibrium condition, where the free energy function is concave, phase separation will occur resulting in the following relationship for the interaction parameter, allowing the free parameter γ to be determined (184) through the best linear for χ.

(5.4)

$$\mathrm{X} = A + \frac{B}{T} = \frac{\frac{1}{N_p}\ln\left(\frac{\phi_h}{\phi_l}\right) + \frac{1}{N_s}\ln\left(\frac{1-\phi_l}{1-\phi_h}\right)}{2(\phi_h - \phi_l)}$$

Here N_1 and N_2 are the number of lattice sites occupied by the protein (100) and solvent (1) respectively. ϕ_h and ϕ_l are the volume fraction of the high- and low-density phases respectively. Phase diagrams are created by determining the coexistence phase locations, where the tangent lines are parallel for two points where the energy is minimized, along the free energy curves for varying temperature.

The critical temperature was obtained by an independent manner through 3D Ising modeling (283). 3D Ising modeling was originally used to explain spin phases formations when magnetic fields are applied to a certain lattice structure. The model has since expanded to include different manners of phase separation including liquid-liquid phase separation. For this modeling I is a sample dependent fitting parameter. For liquid-liquid phase separation, simulation densities which contain a clear high- and low-density

are fit to the following equation. Simulations which resulted in a zero concentration for the low-density phase are excluded.

(5.5)

$$\rho_H - \rho_L = I(T_c - T)^{0.325}$$

Results

BLAST results within the Uniprot database resulting in a list of 3930 proteins after duplications were removed (Figure 6.2). The majority (86.7%) of transcription factors have a maximum of <15 glutamines within a 100 amnio acid region. Larger glutamine numbers appear to cluster around certain values showing a peak at 72 glutamines in a 100 amnio acid region. There are only seven human transcription factors that contain a region between 50 and 65 glutamines. Due to the sporadic clustering above 40%, samples were taken above this region. All sequences are predicted to be disordered by PONDR with disorder increasing with increasing number of Q, except in the case of 54%, which has the maximum average PONDR score (Figure 6.3).

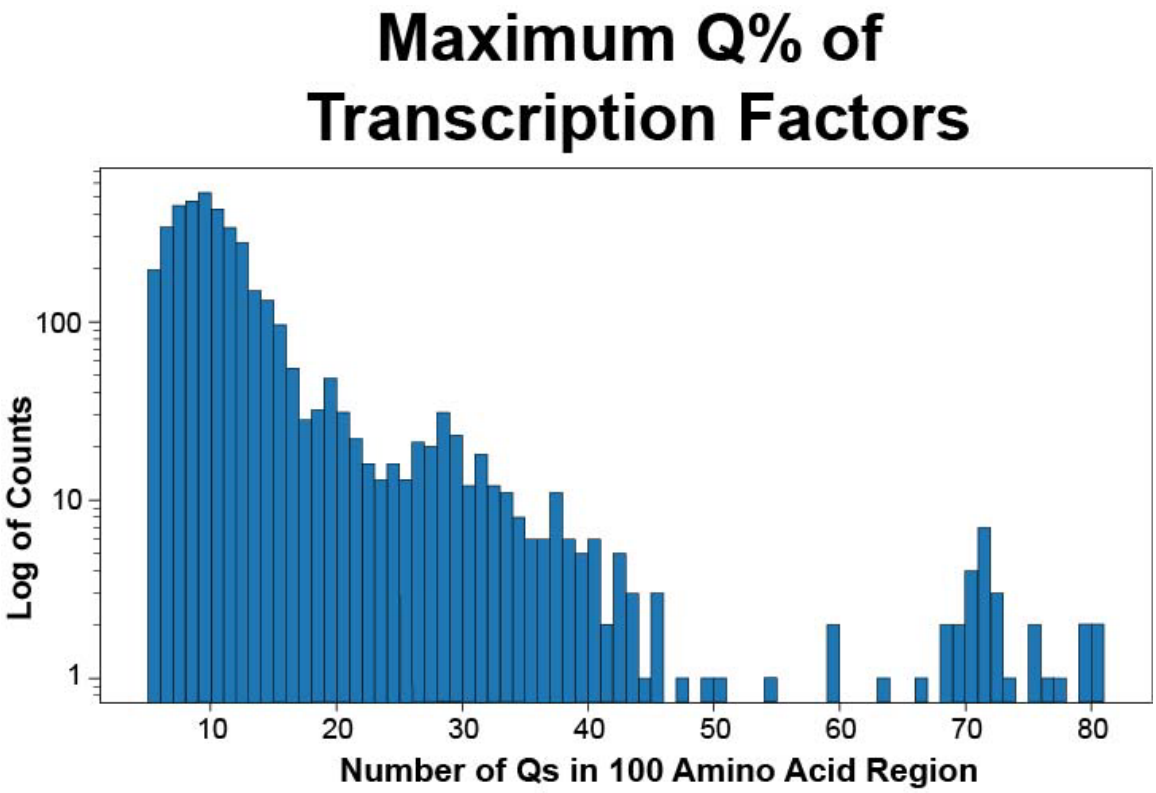


Figure 6.2 **Histogram of glutamine relevance within transcription factors.** The largest number of glutamine residues within a one hundred amino acid was obtained for all human transcription factors with the Uniprot database.

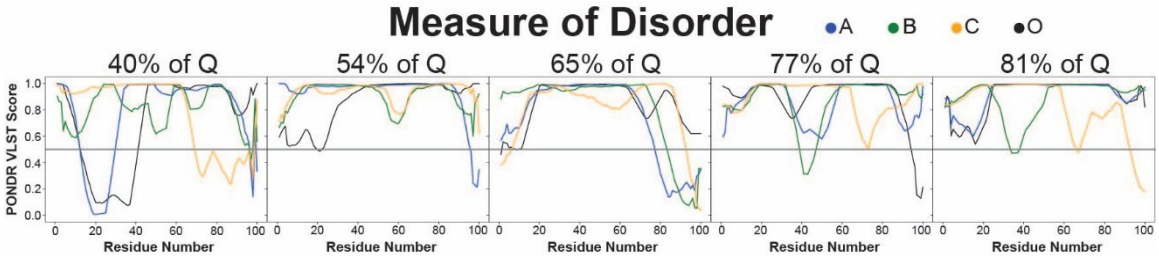


Figure 6.3 **PONDR predictions for the selected sequences.** Panels are ordered, from left to right, in increasing glutamine percentage. The original sequence is shown in black with the sequence scrambles A, B, and C in blue, green, and yellow respectively. A gray line has been added to mark a PONDR score of 0.5, indicating a switch between predicted structed regions below and predicted disordered regions above.

O40				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1062.11	18.06
290	0	0	691.03	8.73
310	0	0	626.48	7.94
330	2.52	1.81	542.67	10.21
340	2.54	3.18	491.34	8.301
350	8.51	2.81	436.13	13.81
500	N/A	N/A	199.28	6.42

A40				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1064.17	47.02
230	0	0	861.41	6.99
250	0	0	805.14	9.78
270	0	0	750.23	9.07
290	0	0	684.49	15.10
300	2.49	3.27	647.34	10.85
305	2.47	2.35	629.25	16.13
310	10.49	8.67	613.68	14.57
315	17.04	8.30	595.17	16.72
320	N/A	N/A	541.30	32.03
330	N/A	N/A	540.84	13.42
350	N/A	N/A	540.93	11.40
500	N/A	N/A	540.93	6.26

B40				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1071.67	29.00
300	0	0	645.22	8.39
310	0	0	612.52	10.61
320	1.64	1.22	565.97	16.35
330	3.03	3.05	526.55	12.71
340	11.72	3.45	468.16	22.45
360	N/A	N/A	290.96	47.49
484	N/A	N/A	290.82	6.86

C40				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1065.88	26.23
290	0	0	672.80	11.54
310	2.86	1.85	595.99	8.51
330	6.03	5.26	501.57	12.58
340	8.68	4.82	423.70	17.11
350	88.38	17.08	354.02	22.86
500	N/A	N/A	220.72	6.49

O54				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1091.45	38.35
290	0	0	714.89	16.00
310	4.17	3.04	645.97	16.48
320	0.23	0.68	602.21	21.48
330	12.09	6.86	562.30	6.052
350	N/A	N/A	408.86	68.70
500	N/A	N/A	408.87	6.87

A54				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1093.53	28.15
290	0	0	697.80	10.11
310	0	0	623.22	7.13
330	1.32	1.39	528.34	13.95
340	5.32	2.66	457.72	29.49
350	78.84	66.42	435.04	15.97
500	N/A	N/A	233.31	5.19

B54				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1086.11	23.23
290	0	0	705.15	8.22
310	1.51	1.17	631.75	6.87
330	4.41	1.63	546.55	11.35
350	71.23	27.35	431.41	27.19
500	N/A	N/A	247.52	7.43

C54				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1095.72	24.22
290	0	0	703.21	9.93
310	1.69	0.95	633.67	8.07
330	6.44	1.46	549.14	11.20
350	12.95	4.13	405.00	42.11
500	N/A	N/A	273.6789	6.52

O65				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1086.94	22.99
290	0	0	699.07	4.72
310	0.76	1.14	627.54	5.14
330	2.72	2.65	543.03	13.02
350	51.41	14.73	425.80	16.37
500	N/A	N/A	237.06	5.65

A65				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1079.26	28.80
290	0	0	686.47	7.30
310	1.73	1.55	615.45	8.35
330	4.90	0.60	523.70	11.97
350	18.87	3.89	360.31	52.72
500	N/A	N/A	252.95	6.78

B65				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1069.53	26.88
290	0	0	688.03	5.38
310	0	0	614.69	10.08
330	4.05	0.81	536.87	8.26
340	4.96	1.68	475.99	13.35
350	11.62	5.68	394.19	28.02
500	N/A	N/A	234.56	5.95

C65				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1081.49	23.55
290	0	0	683.83	7.90
310	4.33	1.39	611.32	8.89
330	8.11	2.34	529.42	10.21
350	139.28	15.66	426.56	17.22
500	N/A	N/A	300.55	8.21

O77				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1062.27	29.07
280	0	0	611.14	11.90
290	0	0	568.45	12.22
300	12.05	4.64	504.41	10.85
305	18.28	3.02	480.85	12.91
310	22.69	11.84	415.17	44.50
315	25.31	6.15	352.10	58.75
320	N/A	N/A	254.86	69.38
330	N/A	N/A	254.85	38.97
350	N/A	N/A	254.84	15.40
500	N/A	N/A	254.85	5.39

B77				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1057.68	17.25
230	0	0	803.73	6.52
250	0	0	731.48	10.12
270	0.94	2.09	650.37	13.56
290	3.18	1.37	547.72	20.01
310	106.77	5.39	428.64	16.89
330	N/A	N/A	245.74	50.53
350	N/A	N/A	245.75	27.91
500	N/A	N/A	245.75	25.16

C77				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1064.65	27.49
250	0	0	741.28	8.56
270	0.55	0.82	657.14	8.69
280	2.08	1.20	613.61	10.48
290	3.76	2.56	557.03	9.80
300	6.05	9.20	509.50	14.58
310	15.65	3.68	439.02	18.07
330	N/A	N/A	231.68	39.88
350	N/A	N/A	231.68	19.51
500	N/A	N/A	231.68	4.66

O81				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1040.29	27.14
210	0	0	822.81	4.72
230	0	0	746.37	6.02
250	0	0	664.21	9.52
260	0	0	608.30	8.68
270	3.14	1.39	557.87	8.35
280	7.63	2.60	491.56	18.92
290	12.176	3.68	427.13	13.76
310	N/A	N/A	236.86	28.89
330	N/A	N/A	236.85	15.81
350	N/A	N/A	236.85	10.70
500	N/A	N/A	236.85	5.31

A81				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1029.31	24.15
190	0	0	884.87	4.89
210	0	0	812.83	5.27
230	0	0	731.38	4.52
250	0	0	642.73	8.10
260	0	0	592.12	6.835
265	3.34	1.18	565.69	9.02
270	6.94	2.94	538.19	13.40
275	3.01	2.95	498.35	15.83
280	24.41	10.76	446.23	26.59
290	80.36	16.70	353.53	43.70
310	N/A	N/A	233.22	19.49
350	N/A	N/A	233.21	9.60
500	N/A	N/A	233.21	5.26

B81				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1022.95	32.61
190	0	0	888.59	8.71
210	0	0	815.81	8.71
230	0	0	740.86	7.94
250	1.34	0.48	649.18	13.54
260	2.05	0.43	602.65	8.23
270	5.39	4.58	543.26	10.37
280	12.61	3.02	468.36	23.82
290	89.10	12.41	404.36	7.87
310	N/A	N/A	264.94	24.12
330	N/A	N/A	264.93	17.54
350	N/A	N/A	264.94	13.85
500	N/A	N/A	264.94	6.20

C81				
Temperature (K)	Dilute Density (mg/mL)	Dilute Density Std (mg/mL)	Condensed Density (mg/mL)	Condensed Density Std (mg/mL)
150	0	0	1033.88	24.75
230	0	0	731.24	6.68
250	0	0	640.02	9.61
270	7.71	1.61	543.91	7.81
275	3.93	2.26	527.62	14.70
280	11.35	6.81	485.26	13.90
285	9.48	2.11	443.32	20.76
290	141.07	16.64	378.86	23.21
310	N/A	N/A	254.47	39.91
330	N/A	N/A	254.47	9.18
350	N/A	N/A	254.47	11.77
500	N/A	N/A	254.46	4.27

Table 6.5 **Density values from simulations**. All simulated data points are supplied. N/A indicates that phase separation did not occur, thus there is only a single density phase.

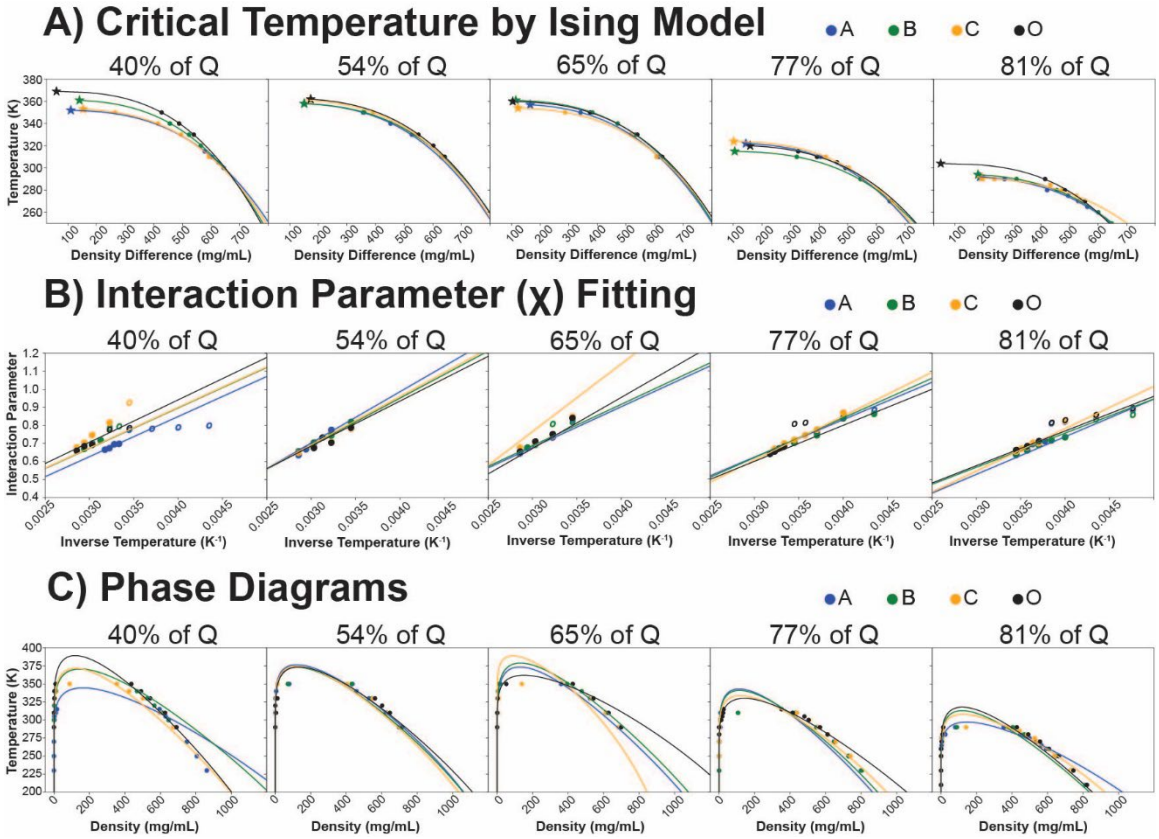


Figure 6.4 Overall fitting results for all samples. Panels are ordered, from left to right, in increasing glutamine percentage. The original sequence is shown in black with the sequence scrambles A, B, and C in blue, green, and yellow respectively. A) Critical Temperature by Ising Modeling. The model is dictated by the solid fit line while data points are marked with closed circles. The predicted critical temperature is marked by a star for each sample. B) Interaction Parameter (χ) Fitting. Linear fits for the interaction parameter (χ) for Flory-Huggins model. The model is dictated by the solid fit line while data points are marked with circles. Closed circles indicate the data points used for fitting, while open circles indicate data points excluded from fitting. C) Phase Diagrams. Phase diagrams constructed by Flory-Huggins modeling. The model is dictated by the solid fit line while data points are marked with closed circles.

Sample with the same percentage of glutamine show similar data fits and critical temperatures for Ising Modeling and phase diagram fitting (Figure 6.4). This indicates scrambling of the sequence has little to no effect upon biomolecular condensate formation. Ising Model shows a good fit using the simulation results which contain a clear high- and low-density phase. Interaction parameter (χ) fitting was more difficult with some samples showing two regions of different linear fits depending upon the concentration of the low-density phase. Unfortunately, at low simulated temperature, a dense droplet form resulting in no diffusion out of the droplet. This results in a low-density phase containing no protein chains. Since a zero value is not an acceptable value for χ fitting, 0.001 mg/mL was assigned for low-density phases with such a condition. This could have led to errors, thus were excluded from χ linear fitting. The resulting phase diagrams show a more rounded top, with higher critical temperatures and smaller low-density phase concentrations than seen in simulations. The fit for phase diagrams leans towards better fitting with the low-density phases than the high-density phases, possibly due to the zero concentration for the low-density phase that occurs at lower temperature.

Sample	Q%	Ising Model T_c (K)	Fitting Param I	Phase Diagram T_c (K)	rho (mg/mL)	dH	dH std	dS	dS std	Disorder
A40	40	352.23	178.63	346.6	1800	222.542	46.416	-0.041	0.151	0.758
B40	40	361.57	168.91	370.6	1636	224.357	44.864	0	0.136	0.810
C40	40	353.66	175.56	371.6	1343	224.897	25.998	0	0.078	0.782
O40	40	369.05	163.94	389.2	1311	235.449	50.629	0	0.149	0.704
A54	54	358.62	176.35	376.6	1330	284.171	37.285	-0.15	0.116	0.931
B54	54	358.59	179.88	373	1362	262.522	20.28	-0.099	0.064	0.909
C54	54	361.91	175.46	358.4	1321	266.634	19.322	-0.107	0.061	0.936
O54	54	362.94	176.92	373.4	1450	249.48	77.239	-0.063	0.248	0.859
A65	65	357.93	174.66	363.4	1398	225.945	26.931	0	0.082	0.739
B65	65	362.08	171.28	370.2	1436	229.285	37.574	0	0.111	0.815
C65	65	354.5	178.27	388.8	980	374.981	65.1	-0.359	0.205	0.807
O65	65	360.37	175.97	362	1658	282.849	35.879	-0.176	0.113	0.844
A77	77	322.53	177.66	340.4	1241	207.478	7.385	0	0.026	0.854
B77	77	315.16	189.25	341	1262	219.67	19.094	-0.039	0.072	0.868
C77	77	324.18	178.24	333.6	1300	243.54	14.408	-0.125	0.051	0.890
O77	77	320.62	185.5	317.4	1570	199.66	10.637	0	0.035	0.888
A81	81	292.96	190.43	297.2	1600	208.602	7.208	-0.097	0.026	0.917
B81	81	294.87	188.95	312.8	1308	189.37	8.702	0	0.032	0.885
C81	81	291.8	208.74	307.8	1349	234.724	25.671	-0.157	0.096	0.822
O81	81	304.02	174.83	317.8	1313	192.274	13.066	0	0.047	0.904

Table 6.6 **Fitting parameters and critical temperatures for 3D Ising modeling and phase diagrams.**

Critical Temperature (K)						
	Ising Modeling		Phase Diagram		Combined	
%Q	Avg	Std	Avg	Std	Avg	Std
40	359.1	6.74	369.5	15.15	364.3	12.82
54	360.5	1.94	370.4	7.04	365.4	7.13
65	358.7	2.85	371.1	10.68	364.9	9.97
77	320.6	3.40	333.1	9.52	326.9	9.49
81	295.9	4.81	308.9	7.62	302.4	9.10

Table 6.7 **Critical Temperature Results.** The critical temperature for each scramble was averaged together to give an average (Avg) critical temperature for Ising modeling and phase diagram fitting. Combined is the combined average for both critical temperature methods. The standard deviation (Std) for each averaging is shown as well. All temperatures are in units of Kelvin.

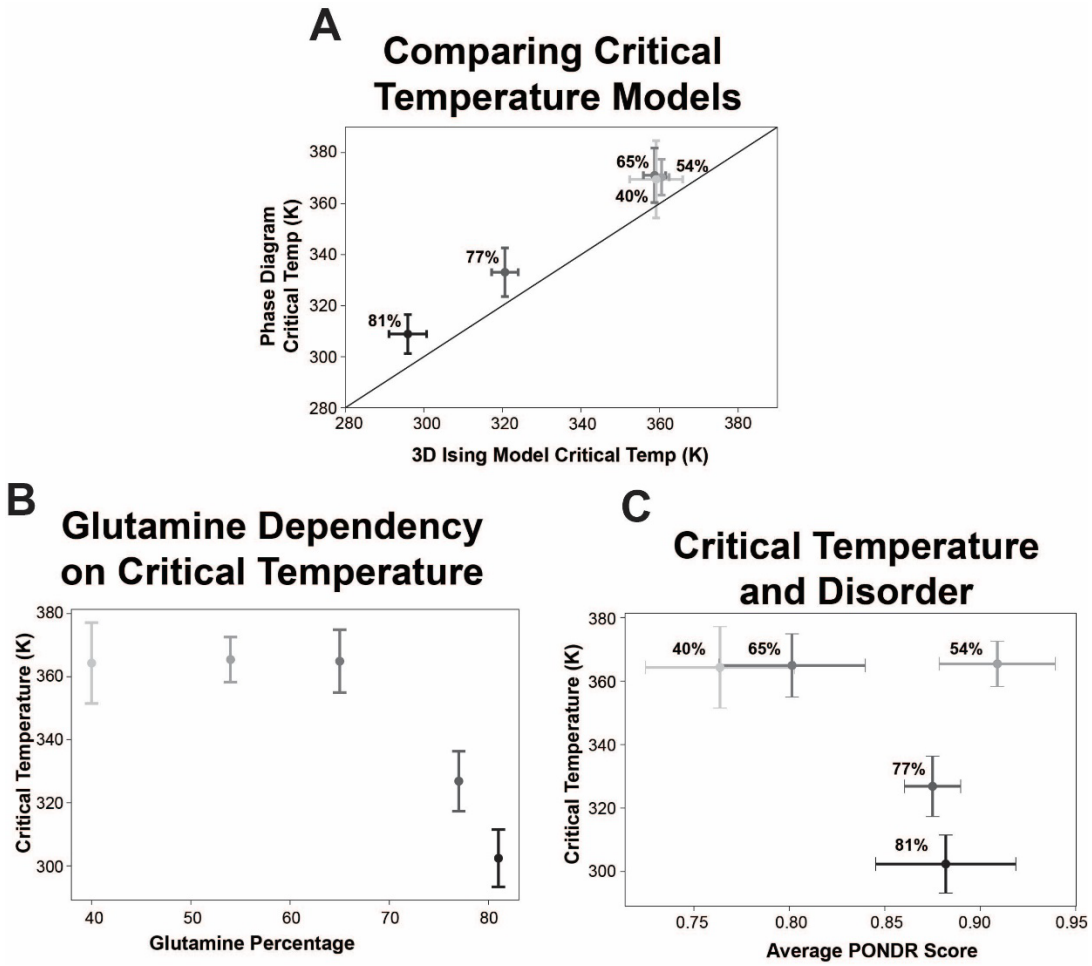


Figure 6.5 **Critical temperature values for Qrich regions.** All error bars denote the standard deviation for averages. A) Comparison of critical temperature values obtained through phase diagram fitting compared to 3D Ising Modeling. Black line is added for matching values comparison. B) Average critical temperature for each glutamine percentage. C) Comparing Average Ponder Score for each glutamine percentage to the average critical temperature value, where the critical temperature is an average combining the Ising Model and Flory-Huggins fitting for the phase diagram.

When comparing the critical temperature obtain through 3D Ising modeling and phase diagram fitting, there is a constant offset with 3D Ising modeling giving lower critical temperature, closer to those seen for single phase temperatures obtained by the simulation (Figure 6.5A). As the number of glutamines is increased, the critical temperature shows a decrease (Figure 6.5B), indicating addition of glutamine into the system decreases biomolecular condensate stability. There is no correlation between critical temperature and predicted disorder by PONDR (Figure 6.5C).

Discussion and Conclusion

Despite the initial assumption that increasing glutamine would promote phase separation thus assisting in biomolecular condensate stability, it is found that higher glutamine percentages result in critical temperature decreases. This could be due to the addition of glutamine displacing the other amino acids that act as stronger stickers for the sticker-spacer model, since glutamine is modeled as a spacer within the simulations. Biomolecular condensate stability is related to the critical temperature due to the relationship between temperature and entropy. An increase in temperature allows the entropy to overcome enthalpy. Thus, lower critical temperatures imply less entropy contributions are required to overcome the enthalpy that is driving phase separation.

The 3D icing modeling shows better critical temperatures that relate to the simulations, while the Floyd Huggins model gives slightly larger critical temperatures than seen in the simulations. While the Floyd Huggins theory does not show accurate phase diagram fitting close to the critical temperature, the simulation result in a flatter top

to the phase diagram than the Flory Huggins fitting. This could be due to the simulation not appropriately capturing the nuances that occur as they approach the critical temperature or Flory Huggins modeling does correctly capture interactions that occur between proteins due to the complexity within intramolecular forces.

By comparing the PONDR predictions for disorder to the obtained critical temperatures no correlation is seen. This implies that disorder is not the main driving force for phase separation. Other factors besides pure disorder must play a role. This agrees with the findings that not all disordered regions will promote phase separation and form biomolecular condensates. Therefore, increasing disorder does not necessarily increase the stability of biomolecular condensates. Comparing the expansion and intramolecular forces of IDPs that participate with biomolecular condensation could be enlightening.

The known aggregation aspect of high glutamine regions is not seen or obtained by the simulations. This implies that the force field for such simulations is not considering aggregation type nature. Due to the lower stability of the liquid phase this implies there might be a lower energy barrier between the liquid to solid transitions for higher glutamine rich regions. However, simulations are not capturing or obtaining this transition. This may be due to the lack of secondary structure considerations when coarse graining by amino acids. Course green simulations which consider the secondary structure might lead to more accurate phase diagram predictions for prion-like domains and help explain how liquid-like states transition to more solid-like aggregates. This could be done by using a more rigorous coarse graining methods that take into

consideration the backbone of the protein as well as the attached functional groups. Further proof is shown with the higher Q-rich percentages. Such amino acid sequences should aggregate such as polyQ regions which lead to Huntington's diseases when at the high concentrations occurring within the simulations. However, no such aggregation is seen within the simulations, with all samples eventually leading to a homogenous state once the critical temperature is reached.

Overall, LAMMPS remains the standard for phase separation predictions, but modifications could be introduced to help understand how secondary structure, thus intramolecular forces, play a role in biomolecular formation. Evidence of this comes from the structural dependence for prions while understanding prion-like domains drive phase separation. Known prion behavior, such as aggregation, is not captured in the LAMMPS simulations.

CHAPTER SEVEN

CONCLUDING REMARKS

Biological systems require signaling to interrupt and react to outside, environmental stimuli. Signaling can take many forms, such as a change in protein configuration due to the presence of a binding ligand or density dependency on temperature as seen with biomolecular condensates. Cell signaling can be very complex, resulting in a cascade of various factors leading to the final cellular response. Obtaining the correct cellular response to outside environmental factors is essential for the continuation of life. All signaling requires a change in the state of the system, otherwise known as dynamics. Without dynamics, life ceases to exist. Dynamics and cell signaling remain at the core of this dissertation.

Proteins have many different manners in which they interact. The C-terminal domain of GluN2B (N2B) appears disordered, due to fast dynamics as it samples numerous conformations. The sampled conformations are driven through self-interactions within N2B, known as intramolecular forces. N2B is thought to signal through glassy transitions states, a phenomenon which is not yet well understood, but of great importance to the biophysics field. Biomolecular condensates also have high dependencies upon environmental factors such as temperature and salt concentration. Biomolecular condensates are formed and stabilized by interactions between different proteins, showing the importance of intermolecular forces for biological systems. Since many biomolecular condensates are found to be formed by contributions of IDPs, the

balance between intra- and intermolecular forces is highly important, though not yet well understood.

While polymer modeling is used throughout biophysics to model biomolecular condensate formation and IDPs it seems to be lacking for both models. Coarse grain polymer models, such as those used in LAMMPS, lack a consideration of secondary structures and protein backbone orientation. Secondary structures are highly important in regard to prion-like diseases and aggregation. Healthy and diseased prion proteins consist of the same amino acid sequence, thus will coarse grain in the same manner. Under the traditional coarse graining force fields, healthy and diseased prions will interact and respond in the same manner. Traditionally IDPs are thought to lack secondary structure, but Discrete Molecular Dynamic (DMD) simulations along with second time scale conformational changes seen in other experiments (264), shows this is not the case. By discarding secondary structure considerations, information is lost, including the importance of intramolecular interactions.

Overall, protein models need to be created that consider both intramolecular and intermolecular interactions to fully understand signaling and dynamics of biological systems. The manner proteins interact with themselves will alter how they interact with the proteins around them and vice versa. Polymer modeling for IDPs needs to be advanced to take such manners into consideration. This can include electrostatic interactions and hydrophobic considerations. Such considerations are currently being applied, but the process is difficult and computationally heavy (284,285). Coarse grain models can drastically improve by increasing the coarse graining technique to include

backbone considerations. Such methods already exist for DNA (189,190,200) and are starting to become easier for proteins (286). While increasing the computational cost by increasing the decrease of freedom within the simulation, such considerations are required to fully understand and probe the dense system of biomolecular condensates. Such advances will also be useful in the field of protein aggregation.

APPENDICES

Appendix A

Increasing Signal Intensity of Fluorescent Oligo-Labeled Antibodies

Increasing Signal Intensity of Fluorescent Oligo-Labeled Antibodies

Madeline E. McCarthy[1], Danielle R. Latham[2], Charlotte V. Haskell[1], Nishi D. Patel[1], Zachariah A. Pittman[1], Hugo Sanabria[2], and Marc R. Birtwistle[1,3,#]

[1]Department of Chemical and Biomolecular Engineering, Clemson University
[2]Department of Physics and Astronomy, Clemson University
[3]Department of Bioengineering, Clemson University
[#]Correspondence: mbirtwi@clemson.edu

Abstract

While full-spectrum flow cytometry has increased antibody-based multiplexing, yet further increases remain potentially impactful. We recently proposed how fluorescence Multiplexing using Spectral Imaging and Combinatorics (MuSIC) could do so using tandem dyes and an oligo-based antibody labeling method. In this work, we found that such labeled antibodies had significantly lower signal intensity than conventionally-labeled antibodies in human cell experiments. To improve signal intensity, we tested moving the fluorophores from the original external (ext.) 5' or 3' end-labeled orientation to internal (int.) fluorophore modifications. Cell-free spectrophotometer measurements showed a ~6-fold signal intensity increase of the new int. configuration compared to the previous ext. configuration. Time-resolved fluorescence spectroscopy and fluorescence correlation spectroscopy showed that ~3-fold brightness difference is due to static quenching. Spectral flow cytometry experiments using peripheral blood mononuclear cells stained with anti-CD8 antibodies showed that int. MuSIC probe-labeled antibodies have signal intensity equal to or greater than conventionally-labeled antibodies with similar estimated proportion of CD8+ lymphocytes. The antibody labeling approach is general and can be broadly applied to many biological and diagnostic applications.

Introduction

Fluorescent antibodies are an integral tool for biological and diagnostic applications[1]. One application for fluorescent antibodies is flow cytometry[2]. The use of fluorescent antibodies with conventional flow cytometers is restricted to typically 3-4 markers, but up to ~10-15 markers have been reported[2–4]. The restriction is largely due to spectral overlap between fluorophores, limiting the number of analytes that can be reliably detected. Regardless, flow cytometry remains a useful platform as it is a cost-effective, high-throughput, and non-destructive method for single-cell analysis[5,6]. Recent advances have led to full-spectrum flow cytometry (FSFC), which captures the entire fluorophore emission spectra, creating a unique spectral fingerprint for each fluorophore[7,8]. This allows fluorophores with similar peak emissions to be used in the same panel, so long as they have distinctive spectral signatures. FSFC has enabled the detection of up to 40 markers simultaneously[9], but further multiplexing capabilities are stunted by the number of commercially available dyes that are compatible in a single panel. Moreover, FSFC is still far from the multiplexing capabilities of methods such as single-cell RNA sequencing, which has the ability to identify 100s-10,000s of markers[10,11].

The 40-plex FSFC panel largely relies on single-dye fluorescent antibodies, with relatively few tandem-dye fluorescent antibodies[9]. We recently developed Multiplexing using Spectral Imaging and Combinatorics (MuSIC), which uses combinations of currently available fluorophores to create spectrally-unique MuSIC probes[12]. MuSIC probe-labeled antibodies may expand the multiplexing capability for FSFC by providing new tandem probes. Previously, we proposed an oligo-based method for covalently

labeling antibodies with MuSIC probes **(Fig 1A-B)** and validated this method using spin column purification, absorbance measurements, and Protein A beads / spectral flow cytometry[13]. However, this method had yet to be tested on human cells.

In the current study, we first applied our previous method to staining human peripheral blood mononuclear cells (PBMC). In doing so, we found significantly lower staining intensity compared to a conventional antibody labeling kit (Biotium Mix-n-Stain). Consequently, we hypothesized that a different oligo-fluorophore arrangement of the MuSIC probes, using internal fluorophore modifications rather than external fluorophore modifications, could increase the fluorescent signal intensity of MuSIC-probe labeled antibodies. Results showed that the new method with internal fluorophore modifications produced ~6-fold increase in fluorescent signal compared to the previous method. Biophysical characterization showed that ~3-fold of this difference is due to fluorescence static quenching. We then compared the internally modified MuSIC-probe labeled antibodies to conventionally labeled antibodies by staining PBMCs. Results showed that the new internal labeling method has ~2-fold increase in fluorescent signal over the conventionally labeled antibodies while having no significant difference in the estimated % of CD8+ lymphocytes. This increased fluorescent signal suggests the potential of MuSIC-probe labeled antibodies to add to the existing capabilities of FCFS, by providing new spectrally unique fluorescent antibodies with comparable intensity. Such antibodies are not restricted to FSFC but could be useful for other biomedical applications such as tissue heterogeneity studies with immunofluorescence imaging when spectral detection is available.

Methods

Measuring fluorescent oligo emission spectra

All oligos (Integrated DNA Technologies, Table 1) are resuspended in ddH20 at 100 μM. In a black 96-well plate (Fisher Scientific Cat: 655900), 200 μmols of the fluorescent oligo(s) is added to the well and the volume is brought up to 50μl with PBS. The fluorescent emission spectra are gathered using a Synergy MX microplate reader (Biotek) with parameters set to a slit width of 9 nm, taking readings from the top, an excitation wavelength set to the maximum excitation wavelength for that fluorophore, and an emission wavelength starting 30 nm after the excitation wavelength (Table 2) and emission collected at every nm. The maximum emission intensity was used to quantify results in Fig. 2B.

Labeling Antibodies

Antibodies are conjugated as previously described[13]. In short, the antibody (CD8 clone RPA-T8; Biolegend Cat: 301002) is incubated with DBCO-Peg5-NHS Ester (linker; 10mM in DMSO; Click Chemistry Tools Cat: 1378531-80-6) in 60 molar excess (10 μg of antibody and 2.8 μg of linker) for 30 minutes at room temperature. Post-incubation, the excess linker is removed with Amicon Ultra 100 kDa molecular weight cut-off filters (Fisher Scientific Cat: UFC5100BK). The antibody-linker retentate is collected. Two oligo complexes are created using external (ext.) or internal (int.) fluorophore modifications.

For externally-modified oligos, a 20 bp oligo with a 5’ fluorophore modification (5’ donor strand) and a 20 bp oligo with a 3’ fluorophore modification (3’ acceptor strand)

are co-hybridized to a 55 bp oligo with a 5' azide modification (docking strand) (Integrated DNA Technologies, Table 1) in a 1:1:1 ratio (0.4 nmol of each oligo) to form the ext. oligo complex.

For internally-modified oligos, a 15bp oligo with a 3' azide modification (azide strand) and a 50 bp oligo with an internal fluorophore modification (int. acceptor strand) are co-hybridized to a 65 bp oligo with an internal fluorophore modification (int. donor strand) (Integrated DNA Technologies, Table 2) at a 1:1:1 ratio to one another (0.4 nmol of each oligo) to form the int. oligo complex.

For each, oligo mixtures are incubated for five minutes at room temperature in the dark to allow for complex formation. These complexes (0.4 nmol of each oligo) are then added to the antibody-linker retentate at a 6-molar excess to the original 10 ug of antibody. The volume is brought up to 100 µl with PBS and incubated at 4°C overnight in the dark.

Conventionally labeled antibodies are labeled as per the manufacturer's instructions (Biotium, Cat: 92446). In short, CD8 antibodies are covalently labeled with CF488A dyes using the Biotium mix-n-stain kit.

Preparing Peripheral Blood Mononuclear Cells

Normal Peripheral Blood Mononuclear Cells (PBMCs) (Precision for Medicine; 10M cells/vial) are thawed and counted with a hemacytometer. Cells are washed twice with cold (4^{o}C) stain buffer (0.01 g/ml BSA in PBS) at 300 x g for 5 min. Post-wash, the cells are resuspended in cold stain buffer and divided into 100 µl aliquots containing 10^{6} cells.

Staining PBMCs

In order to block non-specific Fc-mediated interaction,1 μg of normal Rabbit IgG (ThermoFisher Cat: 31235) is added to the cell sample and incubated for 10 minutes at room temperature. Conventionally, ext., and int. labeled-antibodies are made for staining using the protocols described above (10μg of antibody each); (1) CD8 (clone RPA-T8; Biolegend Cat: 301002) labeled with Atto488 ext. MuSIC probes, (2) CD8 (clone RPA-T8; Biolegend Cat: 301002) labeled with Atto488 int. MuSIC probes, and (3) CD8 (clone RPA-T8; Biolegend Cat: 301002) labeled with CF488A (Biotium Cat: 92446), Antibody concentration is adjusted to 0.25 ug/ul for each sample. The labeled CD8 antibody is added to the cell sample at the appropriate amount as per manufacturer's recommendations (2μg CD8 antibody / 10^6 cells) and allowed to incubate in the dark for 20 minutes on ice. Post-incubation, cells are washed twice with 1 ml of cold staining buffer at 300 x g for 5 min. The final cell pellet is resuspended in 0.5 ml of cold staining buffer.

Flow Cytometry

Stained PBMC samples are analyzed using a Cytek Aurora spectral flow cytometer. First, unstained PBMCs are assayed with the events to record set to 10,000. The SpectroFlo software (Cytek) is used to gate single cells (lymphocytes, monocytes, and granulocytes) by forward and side scatter. We then further gate specifically over the lymphocyte population, as typical based on light scattering distributions[14]. Using these same settings, the stained cell samples are assayed. To compare fluorescence intensity between stained samples we calculate the median intensity of the positively stained

cells in the maximum emission channel (B2) using the Spectroflo software. Positively stained cells are defined as cells with a staining intensity above that of the unstained cell samples using a marker gate.

Size Exclusion Chromatography / Multi-Angle Light Scattering (SEC-MALS)

The purpose of SEC was to purify labeled antibody samples prepared as above to provide monomeric antibody conjugates for biophysical characterization below. The approximate retention behavior and molar mass determination of the SEC column and MALS detector (Agilent, AdvanceBio PL1180-3301) was estimated first using Bovine Serum Albumin at 0.5 mg/mL, 30 μL injection, and a 0.30 mL/min flowrate using a PBS mobile phase. The MALS instrument (Wyatt Technologies, DAWN 785nm) was normalized, aligned, and broadened using the main peak of the eluent BSA, corresponding to unaggregated BSA (~5 min retention time). The online concentration was determined using a refractive index detector (Wyatt Technologies, Optilab WREX-08), and we assigned each sample a *dn/dc* of 0.185. We injected the labeled antibody solutions prepared as described above using the same conditions as the BSA experiment. As the approximate absolute molar mass determination via MALS indicated (**Fig. S1**), the chromatogram showed two distinct regions. Eluent corresponding to the first region between 2.5 and 3.5 minutes elution time had an approximate molar mass range of that expected for antibody-oligo conjugates (with a degree of labeling spectrum approximately between 1 and 6). This fraction was collected into vials and preserved for further analysis.

Fluorescent Correlation Spectroscopy and Time-Resolved Fluorescence and Anisotropy

Freely diffusing samples diluted to sub-nM concentration were analyzed using a custom-built confocal microscope[15]. Samples were excited with a 485 nm pulsed diode laser (LDH-D-C-485, PicoQuant, Germany) operated at 40 MHz. The laser power at the objective was 141 µW. Emission was detected via two PMA detectors (PicoQuant, Germany), allowing for separation into parallel and perpendicular polarization components. A clean-up emission filter (ET525/50, Semrock) is placed before each detector. To ensure temporal data registration of the two synchronized emission channels, we used a HydraHarp 400 TCSPC module (PicoQuant, Germany) in Time-Tagged Time-Resolved mode with a resolution of 1 ps.

Samples were imaged in NUNC chambers (Lab-Tek, Thermo Scientific) that were pre-coated with a solution of 0.01% Tween 20 (Thermo Scientific) in water for 30 min to minimize surface adsorption. Before measurements, chambers were rinsed with buffer to ensure clean measurements. The instrument response function (IRF) was found by measuring water while the protein-free buffer was used for background subtraction. Samples were diluted in charcoal-filtered PBS (10 mM sodium phosphate, pH 7.4, 137 mM NaCl, 2.7 mM KCl) to ~500pM and measured for 2 minutes.

Software correlations were fit with a 3-dimensional Gaussian with two triplet terms (Eq. 1). The confocal geometric parameter (ω_o) was determined using Rhodamine 110 as a standard. Diffusion time (t_{diff}), molecule count (N), baseline term (G_∞), dark state times (t_{T_1} and t_{T_2}), and their corresponding fractions (A_{T_1} and A_{T_2}) were considered as free parameters.

$$G(t_c) = G_\infty + \frac{1}{N} \cdot \frac{1}{1+\frac{t_c}{|t_{diff}|}} \cdot \frac{1}{\sqrt{\left(1+\frac{t_c}{\omega_o \cdot |t_{diff}|}\right)}} \cdot \left(1 - |A_{T_1}| + |A_{T_1}| e^{-\left(\frac{t_c}{|t_{T_1}|}\right)} - |A_{T_2}| + |A_{T_2}| e^{-\left(\frac{t_c}{|t_{T_2}|}\right)}\right) \quad (1)$$

Fluorescence lifetime and rotational correlation times for the samples were found by determining the minimum number of free parameters required to minimize the χ^2. This was determined to be two fluorescence lifetimes and one rotational time for each sample. The comparison between parallel and perpendicular polarized light about the original laser pulse gives the time-resolved fluorescence anisotropy (r(t) – Eq. 2). The time-resolved anisotropy was fit to Eq. 3.

$$r(t) = \frac{I_{parallel}(t) - \left(\frac{1}{G}\right) I_{perpendicular}(t)}{I_{parallel(t)} + \left(\frac{2}{G}\right) I_{perpendicular}(t)} \quad (2)$$

$$r(t) = r_\infty + (r_o - r_\infty) e^{\left(-\frac{t}{\tau_r}\right)} \cong r_o e^{\left(-\frac{t}{\tau_r}\right)} \quad (3)$$

For lifetime fitting, the parallel and perpendicular components were combined using the G-factor (G=1.04) determined using Rhodamine 110 as a standard. Then the fluorescence decays were fit using Eq. 4 with two fluorescence lifetimes for minimizing χ^2. Then Eq. 5 was used to determine the species average lifetime.

$$I(t) = \sum A_n \, e^{-\left(\frac{t}{\tau_n}\right)} \quad (4)$$

$$\langle \tau \rangle_x = \sum A_n \cdot \tau_n \quad (5)$$

Results

We previously developed a method for labeling antibodies with combinations of fluorophores (i.e. MuSIC probes)[13]. In short, an oligo complex containing fluorescent molecules is conjugated to the antibody via a DBCO-Peg5-NHS ester (referred to as the linker) (**Fig 1A**). Here the oligo complex is composed of a 20 bp oligo with a 5' fluorophore modification (referred to as the 5' donor strand) and a 20 bp oligo with a 3' fluorophore modification (referred to as the 3' acceptor strand) that are co-hybridized to a 55bp oligo with a 5' azide modification (referred to as the docking strand) to form the externally labeled (ext.) oligo complex (**Fig 1B**). We previously demonstrated our ability to covalently label antibodies with MuSIC probes using this method and validated the labeling protocol with spin-column purification, absorbance measurements, and FSFC measurements with protein A beads bound to (i) Cy3, (ii) Tex615, and (iii) Cy3-Tex615 ext. oligo labeled antibodies[13].

Since this method had only been validated using beads, we asked whether this method would work when staining peripheral blood mononuclear cells (PBMCs)? We created an ext. oligo complex using an Atto488 5' donor strand and an Atto488 3' acceptor strand as the MuSIC probe and conjugated it to anti-CD8 antibodies. For comparison, we used a commercially available Biotium Mix-n-Stain kit to conventionally label CD8 antibodies with CF488A dye, which is reported to have comparable fluorescent properties (excitation peak, emission peak, and brightness) to Atto488[16]. PBMCs were stained with each antibody batch and analyzed by FSFC. Results showed that the median signal intensity of cells stained with the ext. labeled MuSIC probe was

~2-fold (p-value=0.01) lower compared to cells stained with conventionally labeled antibodies (**Fig 1C**).

We then asked how we can increase the signal intensity of MuSIC probe-labeled antibodies. We reasoned that the lower fluorescence signal was not due to the degree of labeling because it was previously calculated to be within the standard range[13,17]. Some degree of difference in signal intensity may be due to differences in dye properties between CF488A and Atto488, although as mentioned above, the dyes have similar characteristics. To investigate whether the docking strand and/or hybridization played a role, we examined the fluorescence emission intensity of Atto488 5' donor strands and Atto488 3' acceptor strands alone in solution and when co-hybridized to the docking strand (**Fig 2A**). We found that the hybridization of the 5' donor and 3' acceptor strands to the docking strand results in a significant decrease in fluorescent signal, as compared to the strands on their own.

We further wondered whether this was a fluorophore-specific phenomenon or if it occurred for other fluorophores. Therefore, we compared the emission intensity with and without docking strand for 15 different fluorophore-conjugated 5' donor strands and 3' acceptor strands (**Fig 2B**). Signal decreased with docking strand for all but five of the fluorophore-conjugated strands that were tested. Previous studies showed that there can be a significant change in fluorescence when oligo-strands containing an end-fluorophore modification are hybridized to strands containing an overhang[18], such as in our ext. oligo complex.

These findings led us to hypothesize that if the fluorophores within the oligo complex had a different orientation, it could give an increased fluorescent signal. To test

this, we adjusted the configuration of the ext. oligo complex to contain oligos with internally (int) conjugated fluorophores. This new oligo complex consists of the 50 bp int. acceptor strand and a 15 bp azide strand which both co-hybridize to the 65 bp int. donor strand (**Fig 3A**). The purpose of a separate azide strand is to reduce the cost of oligo production, due to the increased difficulty of synthesizing an oligo with two modifications. The new donor and acceptor strands both have an internal fluorophore modification (int donor and int acceptor), rather than 5' and 3' end fluorophore modification, respectively. We then created int. and ext. oligo complexes (both using Atto488 conjugated strands) and measured their fluorescent emission spectra. We observed a ~6-fold fluorescent signal increase of the int. oligo complex compared to the ext. oligo complex in solution (**Fig 3B**).

To understand the fluorescent signal differences between the int. oligo complex and ext. oligo complex, we used Time Resolve Fluorescence Spectroscopy (TRFs) and Fluorescence Correlation Spectroscopy (FCS) (**Fig 4**). We found that the ext. oligo complex undergoes more dynamic quenching than the int. oligo complex, as shown by the differences in the fluorescence decay, which can also be quantified using the species average lifetime (**Fig 4A, Table 3**). The ext. oligo complex also spends more time in the dark triplet state than the int. oligo complex, as shown by the differences in the correlation curves (**Fig 4B, Table 3**). Further, since FCS only monitors bright molecules, a single (bright) int. oligo complex is only 1.6 times brighter than an ext. oligo complex (**Table 3**). Considering the observed ~6-fold difference in intensity (**Fig. 3B)** but < 2-fold difference in the molecular brightness (**Table 3**) of the int. oligo complex relative to the ext. oligo complex, we conclude that ~3-fold difference in intensity is due

to static quenching in the ext. oligo complex. As expected, both oligo complexes show similar diffusion and rotational correlation times **(Fig 4B and C)**.

With this increase in signal intensity, we then asked how new int. MuSIC probe-labeled antibodies would compare to conventionally labeled antibodies when staining PBMCs for estimation of specific cell type abundances. Similar to above, int. oligo complexes with Atto488 were conjugated to CD8 antibodies to create int. MuSIC probe-labeled antibodies and CF488A was conjugated to CD8 antibodies using a Mix-n-stain kit to create the conventionally labeled antibodies. PBMCs were stained with each antibody batch and analyzed by FSFC. The signal intensity of cells stained with the int. labeled MuSIC probe was ~2 fold (p-value=0.03) higher compared to cells stained with conventionally labeled antibodies (**Fig 5A**). When comparing the % of CD8+ lymphocytes detected, we found no significant difference between the int. MuSIC probe-labeled antibodies and conventionally labeled antibodies (**Fig 5B**). These results demonstrate that we were able to improve the design of MuSIC-probe labeled antibodies to increase the signal-to-noise ratio, with staining behavior comparable to conventionally labeled antibodies.

Discussion

Here, we established a method to conjugate two fluorophores to an antibody and stain human cells with an increased signal intensity compared to our previous method and conventionally labeled antibodies, with accurate detection of % of CD8+ lymphocytes. This method builds on our previously established labeling protocol but introduces key modifications to the oligo-fluorophore arrangement of the MuSIC probe. By re-arranging the oligo complex to eliminate the use of 3' or 5' end fluorophore modifications, we observe a significant increase in fluorescent signal for Atto488. Given the prevalence of docking strand-induced signal decrease, we expect this result may often apply to other fluorophores. We used Time Resolve Fluorescence Spectroscopy and Fluorescence Correlation Spectroscopy to compare the old ext. oligo complex to the new int. oligo complex, finding that increased dynamic quenching and time in the dark triplet state explains the decreased fluorescence intensity of the ext. oligo complex. Using the new design, we stained human PBMCs and compared the signal intensity to that of conventionally labeled fluorescent antibodies, and observed a statistically significant increase in the resulting fluorescent signal without creating any significant differences in the % of CD8+ lymphocytes.

To maximize the potential of this new increased intensity probe design, the next step will be to select different combinations of fluorophores to assemble a palette of spectrally unique antibody-conjugated MuSIC probes. Approaches to do so can include stimulation studies for compatibility using a workflow similar to that described in our previous work[19], and then testing the highest-ranked fluorophore combinations experimentally. For these simulations, the emission spectra of each possible MuSIC

probe is generated, and lists of MuSIC probes that are likely to be deconvolvable in a mixture are generated and ranked. This ranking provides prioritization for testing experimentally by measuring the emission spectra of mixtures of MuSIC probes and unmixing them to determine which MuSIC probes can be accurately demultiplexed.

One major application of using MuSIC probe-labeled antibodies with FSFC can be cell-type profiling, which is the process by which a complex mixture of cell types, for example, from blood or tumors, are classified into the fractional composition of its components (e.g., neutrophils, natural killer cells, various types of T and B cells, etc.), based on classification of expression patterns (e.g., CD3 expressed or not)[20]. While there are 40 FSFC dyes available, very few of them are tandem dyes that can be used as uniquely identifiable markers, which limits the number of individual analytes that can be classified simultaneously. However, MuSIC probe-labeled antibodies could be used to expand the number of markers that can be detected by creating new combination fluorophore probes from the current dyes, to enhance current cell-type profiling efforts. FSFC has been previously paired with cell-type profiling to investigate the correlation between CD38 expression in macrophages and the predicted immune response to immune-checkpoint blockade therapy for hepatocellular carcinoma[21]. With a larger palette of compatible fluorescent tags, cell-type profiling efforts could expand further to look at an increased number of cell-type markers, for a more comprehensive view of a patient's immune response to various treatments, or to complement other single cell profiling efforts[22–24].

Additionally, MuSIC probe-labeled antibodies can be applied to a broad range of biological and diagnostic applications that involve the detection of protein expression.

One of these applications can be for tissue imaging, such as recent highly multiplexed efforts[25–30]. If MuSIC probe-labeled antibodies can be combined with spectral imaging, this could allow for highly multiplexed, quantitative tissue imaging. One potential application is cancer, where increasing multiplexing capabilities could improve diagnostic potential by allowing for more tumor markers to be analyzed, thus leading to an increased mapping of tumor heterogeneity[31]. This could impact tumor detection, diagnosis, and treatment.

Although here we focused on increasing the fluorescent signal of oligo-based probes, by titrating the fluorescent oligos, we can decrease the fluorescent signal to a desired level in a controllable manner. Tunable fluorescence intensity is useful; for example, in static light scattering experiments[32], where the sensitive photodiode detectors are easily saturated. Here, they labeled BSA at varying concentrations of fluorescent oligos between 0.03 - 0.10 µM that fluoresced below the saturation limit of the detectors while still achieving desired intensity. Conventional labeling kits would have been too intense, and as most are single reaction use, they can be difficult to control compared to the reported oligo-based probes which offer the unique advantage of reduced, tailorable intensities. In their case, the intensity tunability of the probes enabled a more flexible experimental design capable of separating simultaneous fluorescence and light scattering signals. The tunability of oligo-based fluorescent probe intensity could also be beneficial for cell staining, where some epitopes may have such a high abundance that a reduced fluorescent signal is necessary.

In addition to tunability for probe fluorescence intensity, the new int. oligo arrangement of these probes offers modulation of Förster Resonance Energy Transfer

(FRET) between fluorophore combinations on the donor and acceptor strands. By adjusting the distance (bp) between the two fluorophores, one can increase or decrease the FRET efficiency. By adjusting the FRET efficiency of each combination, there is the potential to increase the number of possible probes even further by creating linearly independent combinations.

We conclude that by using an oligo-based approach with internally-labeled fluorophores, we can increase the signal intensity of MuSIC-probe labeled antibodies. MuSIC probe-labeled antibodies may prove useful to increase multiplexing capabilities of full spectrum flow cytometry, and also more broadly where increased multiplexing at single-cell or sub-cellular resolution is needed, including cell-type profiling, tissue studies, and immunofluorescence imaging.

Acknowledgments

MRB acknowledges funding from Clemson University Creative Inquiry, NIH/NCI Grant R21CA196418, and NIH/NIGMS Grant R35GM141891. MEM received funding from the Department of Education Grant P200A180076. HS acknowledges support from NIH (R01MH081923, 1P20GM130451, R15CA280699) and NSF (CAREER MCB 1749778).

References

(1) Brummelman, J.; Haftmann, C.; Núñez, N. G.; Alvisi, G.; Mazza, E. M. C.; Becher, B.; Lugli, E. Development, Application and Computational Analysis of High-Dimensional Fluorescent Antibody Panels for Single-Cell Flow Cytometry. *Nat. Protoc.* **2019**, *14* (7), 1946–1969. https://doi.org/10.1038/s41596-019-0166-2.

(2) Adan, A.; Alizada, G.; Kiraz, Y.; Baran, Y.; Nalbant, A. Flow Cytometry: Basic Principles and Applications. *Crit. Rev. Biotechnol.* **2017**, *37* (2), 163–176. https://doi.org/10.3109/07388551.2015.1128876.

(3) Newell, E. W.; Sigal, N.; Bendall, S. C.; Nolan, G. P.; Davis, M. M. Cytometry by Time-of-Flight Shows Combinatorial Cytokine Expression and Virus-Specific Cell Niches within a Continuum of CD8+ T Cell Phenotypes. *Immunity* **2012**, *36* (1), 142–152. https://doi.org/10.1016/j.immuni.2012.01.002.

(4) McKinnon, K. M. Multiparameter Conventional Flow Cytometry. In *Flow Cytometry Protocols*; Hawley, T. S., Hawley, R. G., Eds.; Springer New York: New York, NY, 2018; pp 139–150. https://doi.org/10.1007/978-1-4939-7346-0_8.

(5) Niewold, P.; Ashhurst, T. M.; Smith, A. L.; King, N. J. C. Evaluating Spectral Cytometry for Immune Profiling in Viral Disease. *Cytometry A* **2020**, *97* (11), 1165–1179. https://doi.org/10.1002/cyto.a.24211.

(6) McKinnon, K. M. Flow Cytometry: An Overview. *Curr. Protoc. Immunol.* **2018**, *120* (1), 5.1.1-5.1.11. https://doi.org/10.1002/cpim.40.

(7) Jaimes, M. C.; Leipold, M.; Kraker, G.; Amir, E.; Maecker, H.; Lannigan, J. Full Spectrum Flow Cytometry and Mass Cytometry: A 32-Marker Panel Comparison. *Cytometry A* **2022**, *101* (11), 942–959. https://doi.org/10.1002/cyto.a.24565.

(8) Bonilla, D. L.; Reinin, G.; Chua, E. Full Spectrum Flow Cytometry as a Powerful Technology for Cancer Immunotherapy Research. *Front. Mol. Biosci.* **2021**, *7*.

(9) Park, L. M.; Lannigan, J.; Jaimes, M. C. OMIP-069: Forty-Color Full Spectrum Flow Cytometry Panel for Deep Immunophenotyping of Major Cell Subsets in Human Peripheral Blood. *Cytometry A* **2020**, *97* (10), 1044–1051. https://doi.org/10.1002/cyto.a.24213.

(10) Papalexi, E.; Satija, R. Single-Cell RNA Sequencing to Explore Immune Cell Heterogeneity. *Nat. Rev. Immunol.* **2018**, *18* (1), 35–45. https://doi.org/10.1038/nri.2017.76.

(11) Stuart, T.; Satija, R. Integrative Single-Cell Analysis. *Nat. Rev. Genet.* **2019**, *20* (5), 257–272. https://doi.org/10.1038/s41576-019-0093-7.

(12) Holzapfel, H. Y.; Stern, A. D.; Bouhaddou, M.; Anglin, C. M.; Putur, D.; Comer, S.; Birtwistle, M. R. Fluorescence Multiplexing with Spectral Imaging and Combinatorics. *ACS Comb. Sci.* **2018**, *20* (11), 653–659. https://doi.org/10.1021/acscombsci.8b00101.

(13) McCarthy, M. E.; Anglin, C. M.; Peer, H. A.; Boleman, S. A.; Klaubert, S. R.; Birtwistle, M. R. Protocol for Creating Antibodies with Complex Fluorescence Spectra. *Bioconjug. Chem.* **2021**, *32* (6), 1156–1166. https://doi.org/10.1021/acs.bioconjchem.1c00220.

(14) Janols, H.; Bredberg, A.; Thuvesson, I.; Janciauskiene, S.; Grip, O.; Wullt, M. Lymphocyte and Monocyte Flow Cytometry Immunophenotyping as a Diagnostic

Tool in Uncharacteristic Inflammatory Disorders. *BMC Infect. Dis.* **2010**, *10* (1), 205. https://doi.org/10.1186/1471-2334-10-205.

(15) Ma, J.; Yanez-Orozco, I. S.; Rezaei Adariani, S.; Dolino, D.; Jayaraman, V.; Sanabria, H. High Precision FRET at Single-Molecule Level for Biomolecule Structure Determination. *J. Vis. Exp.* **2017**, No. 123, 55623. https://doi.org/10.3791/55623.

(16) Biotium, Inc. CF Dyes, 2020. https://biotium.com/wp-content/uploads/2013/07/CF-Dye-Brochure.pdf.

(17) Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Mao, F.; Leung, W.-Y.; Haugland, R. P. Alexa Dyes, a Series of New Fluorescent Dyes That Yield Exceptionally Bright, Photostable Conjugates. *J. Histochem. Cytochem.* **1999**, *47* (9), 1179–1188. https://doi.org/10.1177/002215549904700910.

(18) Nazarenko, I.; Pires, R.; Lowe, B.; Obaidy, M.; Rashtchian, A. Effect of Primary and Secondary Structure of Oligodeoxyribonucleotides on the Fluorescent Properties of Conjugated Dyes. *Nucleic Acids Res.* **2002**, *30* (9), 2089–2195. https://doi.org/10.1093/nar/30.9.2089.

(19) McCarthy, M. E.; Dodd, W. B.; Lu, X.; Patel, N. D.; Haskell, C. V.; Sanabria, H.; Blenner, M. A.; Birtwistle, M. R. A Theory for High-Throughput Genetic Interaction Screening. bioRxiv October 5, 2022, p 2022.10.05.510977. https://doi.org/10.1101/2022.10.05.510977.

(20) Ermann, J.; Rao, D. A.; Teslovich, N. C.; Brenner, M. B.; Raychaudhuri, S. Immune Cell Profiling to Guide Therapeutic Decisions in Rheumatic Diseases. *Nat. Rev. Rheumatol.* **2015**, *11* (9), 541–551. https://doi.org/10.1038/nrrheum.2015.71.

(21) Ng, H. H. M.; Lee, R. Y.; Goh, S.; Tay, I. S. Y.; Lim, X.; Lee, B.; Chew, V.; Li, H.; Tan, B.; Lim, S.; Lim, J. C. T.; Au, B.; Loh, J. J. H.; Saraf, S.; Connolly, J. E.; Loh, T.; Leow, W. Q.; Lee, J. J. X.; Toh, H. C.; Malavasi, F.; Lee, S. Y.; Chow, P.; Newell, E. W.; Choo, S. P.; Tai, D.; Yeong, J.; Lim, T. K. H. Immunohistochemical Scoring of CD38 in the Tumor Microenvironment Predicts Responsiveness to Anti-PD-1/PD-L1 Immunotherapy in Hepatocellular Carcinoma. *J. Immunother. Cancer* **2020**, *8* (2), e000987. https://doi.org/10.1136/jitc-2020-000987.

(22) Cao, J.; O'Day, D. R.; Pliner, H. A.; Kingsley, P. D.; Deng, M.; Daza, R. M.; Zager, M. A.; Aldinger, K. A.; Blecher-Gonen, R.; Zhang, F.; Spielmann, M.; Palis, J.; Doherty, D.; Steemers, F. J.; Glass, I. A.; Trapnell, C.; Shendure, J. A Human Cell Atlas of Fetal Gene Expression. *Science* **2020**, *370* (6518), eaba7721. https://doi.org/10.1126/science.aba7721.

(23) Haniffa, M.; Taylor, D.; Linnarsson, S.; Aronow, B. J.; Bader, G. D.; Barker, R. A.; Camara, P. G.; Camp, J. G.; Chédotal, A.; Copp, A.; Etchevers, H. C.; Giacobini, P.; Göttgens, B.; Guo, G.; Hupalowska, A.; James, K. R.; Kirby, E.; Kriegstein, A.; Lundeberg, J.; Marioni, J. C.; Meyer, K. B.; Niakan, K. K.; Nilsson, M.; Olabi, B.; Pe'er, D.; Regev, A.; Rood, J.; Rozenblatt-Rosen, O.; Satija, R.; Teichmann, S. A.; Treutlein, B.; Vento-Tormo, R.; Webb, S. A Roadmap for the Human Developmental Cell Atlas. *Nature* **2021**, *597* (7875), 196–205. https://doi.org/10.1038/s41586-021-03620-1.

(24) Human Cell Atlas Standards and Technology Working Group; Rozenblatt-Rosen, O.; Shin, J. W.; Rood, J. E.; Hupalowska, A.; Regev, A.; Heyn, H. Building a High-

Quality Human Cell Atlas. *Nat. Biotechnol.* **2021**, *39* (2), 149–153. https://doi.org/10.1038/s41587-020-00812-4.

(25) Lin, J.-R. R.; Fallahi-Sichani, M.; Sorger, P. K. Highly Multiplexed Imaging of Single Cells Using a High-Throughput Cyclic Immunofluorescence Method. *Nat. Commun.* **2015**, *6*, 8390. https://doi.org/10.1038/ncomms9390.

(26) Lin, J.-R.; Chen, Y.-A.; Campton, D.; Cooper, J.; Coy, S.; Yapp, C.; Tefft, J. B.; McCarty, E.; Ligon, K. L.; Rodig, S. J.; Reese, S.; George, T.; Santagata, S.; Sorger, P. K. High-Plex Immunofluorescence Imaging and Traditional Histology of the Same Tissue Section for Discovering Image-Based Biomarkers. *Nat. Cancer* **2023**, 1–17. https://doi.org/10.1038/s43018-023-00576-1.

(27) Goltsev, Y.; Samusik, N.; Kennedy-Darling, J.; Bhate, S.; Hale, M.; Vazquez, G.; Black, S.; Nolan, G. P. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **2018**, *174* (4), 968-981.e15. https://doi.org/10.1016/j.cell.2018.07.010.

(28) Angelo, M.; Bendall, S. C.; Finck, R.; Hale, M. B.; Hitzman, C.; Borowsky, A. D.; Levenson, R. M.; Lowe, J. B.; Liu, S. D.; Natkunam, Y.; Nolan, G. P. Multiplexed Ion Beam Imaing (MIBI) of Human Breast Tumors. *Nat. Med.* **2014**, *20* (4), 436–442. https://doi.org/10.1038/nm.3488.Multiplexed.

(29) Gut, G.; Herrmann, M. D.; Pelkmans, L. Multiplexed Protein Maps Link Subcellular Organization to Cellular States. *Science* **2018**, *361* (6401). https://doi.org/10.1126/science.aar7042.

(30) Giesen, C.; Wang, H. A. O.; Schapiro, D.; Zivanovic, N.; Jacobs, A.; Hattendorf, B.; Schüffler, P. J.; Grolimund, D.; Buhmann, J. M.; Brandt, S.; Varga, Z.; Wild, P. J.; Günther, D.; Bodenmiller, B. Highly Multiplexed Imaging of Tumor Tissues with Subcellular Resolution by Mass Cytometry. *Nat. Methods* **2014**, *11* (4), 417–422. https://doi.org/10.1038/nmeth.2869.

(31) Andreou, C.; Weissleder, R.; Kircher, M. F. Multiplexed Imaging in Oncology. *Nat. Biomed. Eng.* **2022**, *6* (5), 527–540. https://doi.org/10.1038/s41551-022-00891-5.

(32) Pittman, Z. A.; McCarthy, M. E.; Birtwistle, M. R.; Kitchens, C. L. Method for Improved Fluorescence Corrections for Molar Mass Characterization by Multiangle Light Scattering. *Biomacromolecules* **2022**, *23* (9), 3743–3751. https://doi.org/10.1021/acs.biomac.2c00600.

Tables

Table 1: Sequences for the ext. and int. oligo complexes.

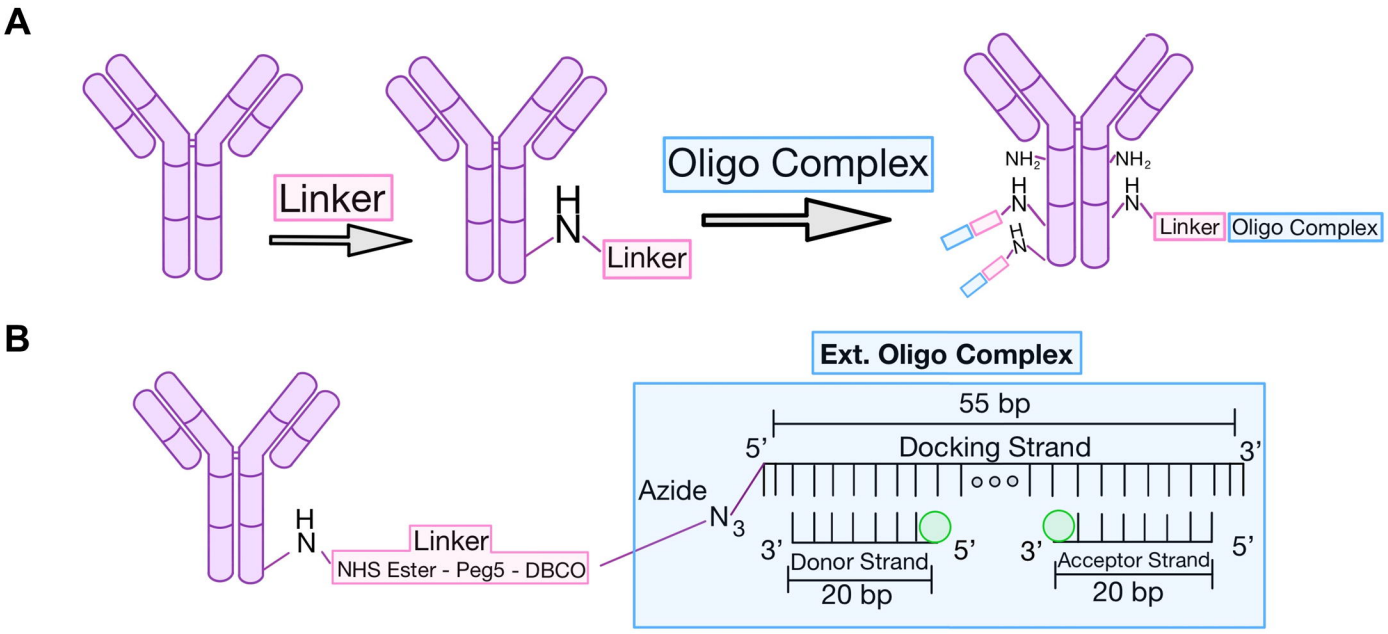
	Component	**Sequence**
Ext. Oligo Complex	Docking Strand	5′-azide-GTG TAG TTC AGG TCA AGA CAT CGT GCG ACC AGT CAG CAT GAG ACT CAT TGG TGC G-3′
	5’ Donor Strand	3′-C AAG TCC AGT TCT GTA GCA C-fluorophore-5′
	3’ Acceptor Strand	3′-fluorophore-CA GTC GTA CTC TGA GTA AC-5′
Int. Oligo Complex	Azide Strand	3’-azide-CGT TAT GAA CCT GA 5’
	Int. Donor Strand	5’ GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT TCT TGT T-fluorophore-A CAA CTC CGA AAT AGG CCG 3’
	Int. Acceptor Strand	3’ CAG ATC CGC TTG CAA ATT CCG C-fluorophore-A GAG ACA AAT GTT GAG GCT TTA TCC GGC 5’

Table 2: Fluorophore modifications for donor and acceptor strands with their corresponding excitation wavelength and emission start wavelength.

Fluorophore Modification	**Excitation (nm)**	**Emission (nm)**
6-FAM (Fluorescein)	490	520
Atto 488	492	522
Atto 532	524	554
MAX (NHS Ester)	527	557
Cy3	534	564
Atto 550	545	575
Tamra (NHS Ester)	553	583
Atto 565	561	591
ROX (NHS Ester)	578	608
TEX 615	583	613
Atto 590	594	624
Atto 633	623	653
Atto 647	632	662
Cy5	638	668
Cy5.5	676	706

Table 3: Fluorescence Correlation Spectroscopy, Time-Resolved Fluorescence and Anisotropy Parameters

	Atto488 Ext. Oligo Complex	Atto488 Int. Oligo Complex
Overall Results		
Molecular Brightness (kHz/molecule)	8.1	13.5
Species Average Lifetime ($\langle\tau\rangle_x$) (ns)	2.58	3.05
Diffusion coefficient ($\mu m^2/s$)	100	120
Quantum Yield	50%	60%
Time Resolved Fluorescence and Anisotropy Parameters		
A_1	57%	68%
t_1 (ns)	4.0	4.0
A_2	43%	32%
t_2 (ns)	0.76	0.92
χ^2	1.65	1.64
Rotational Time (ps)	280	350
Fluorescence Correlation Spectroscopy		
G_0	1.00	1.00
N	0.49	0.50
t_{diff} (ms)	0.71	0.59
ω_o	4.73	4.73
A_{T_1}	0.28	0.26
t_{T_1} (µs)	2.4	0.9
A_{T_2}	0.13	0.082
t_{T_2} (ms)	0.19	0.088
χ^2	0.96	0.86



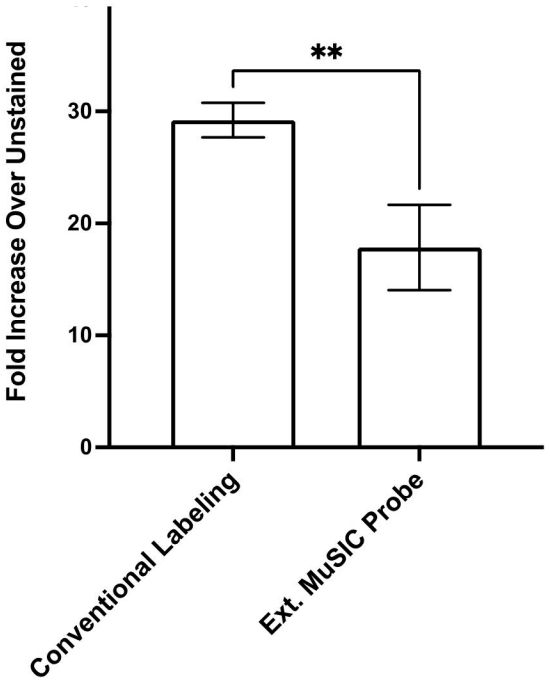


Figure 1. Oligo-based MuSIC probe labeling of antibodies. (A-B) Graphic depicting MuSIC probe labeling. By reacting the NHS ester of the linker with free amines on the antibody, the linker is attached. Fluorophore-labeled donor and acceptor strands are annealed onto the docking strand to form the oligo complex. The azide on the docking strand, in the oligo complex, is reacted with the free DBCO group on the linker to covalently bind the oligo complex to the antibody. There are multiple free amines on each antibody, allowing for the linker to attach at multiple sites, increasing the degree of labeling. (C) Comparison of fluorescence intensity of PBMCs stained with CF488A conventional labeling kit vs Atto488 Ext. MuSIC probes. The fold increase in intensity over unstained cells is significantly greater for the conventional labeling kit compared to the ext. MuSIC probe. Error bars are standard error from triplicate measurements.

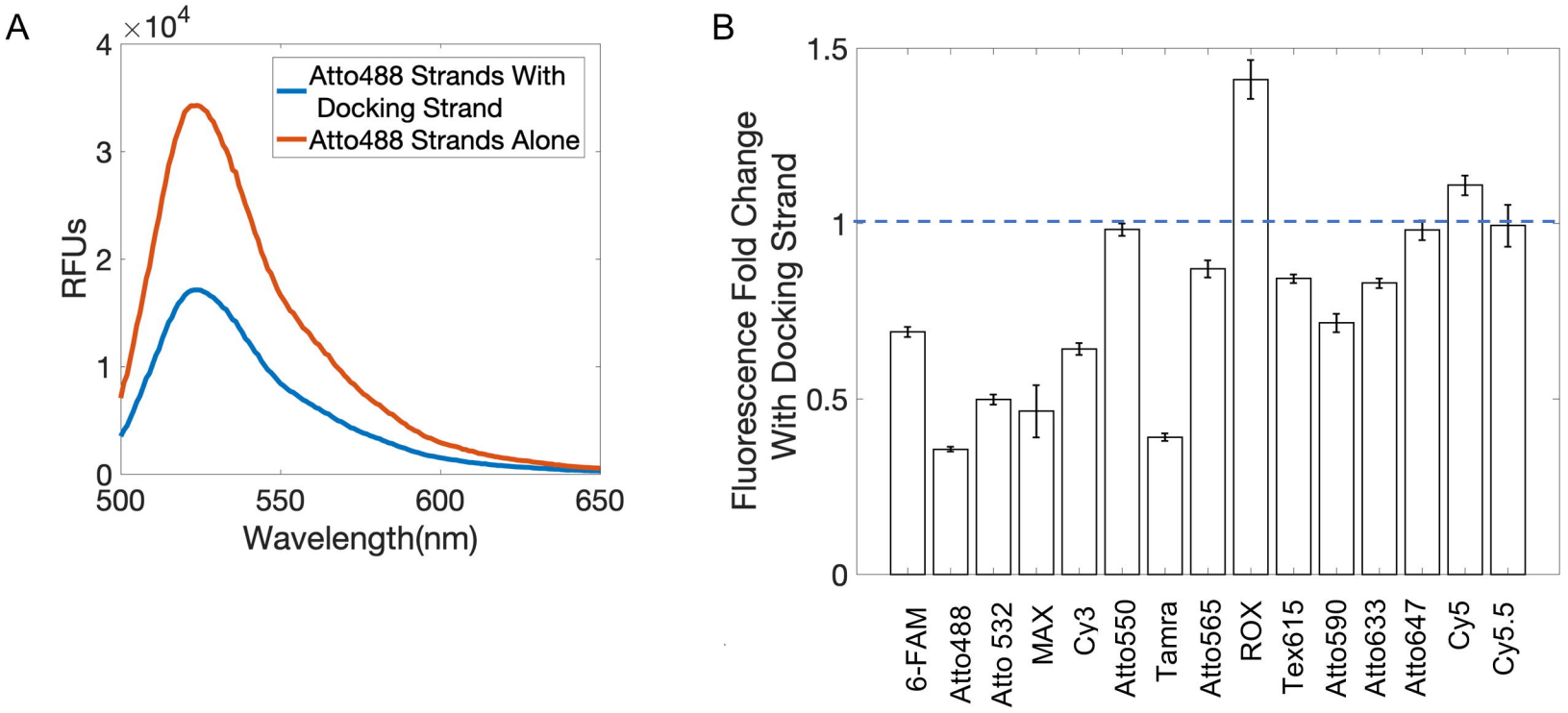


Figure 2. Fluorescence signal change from Docking Strand. (A) Comparison of fluorescence emission spectra, excited at 470nm, of the Atto488 5' Donor and 3' Acceptor strands hybridized to the Docking Strand and when alone in solution with and without the Docking Strand. Data are representative from triplicates. (B) Change in fluorescence intensity of 15 fluorescent oligos when hybridized to the Docking Strand. Error bars are standard error from triplicates.

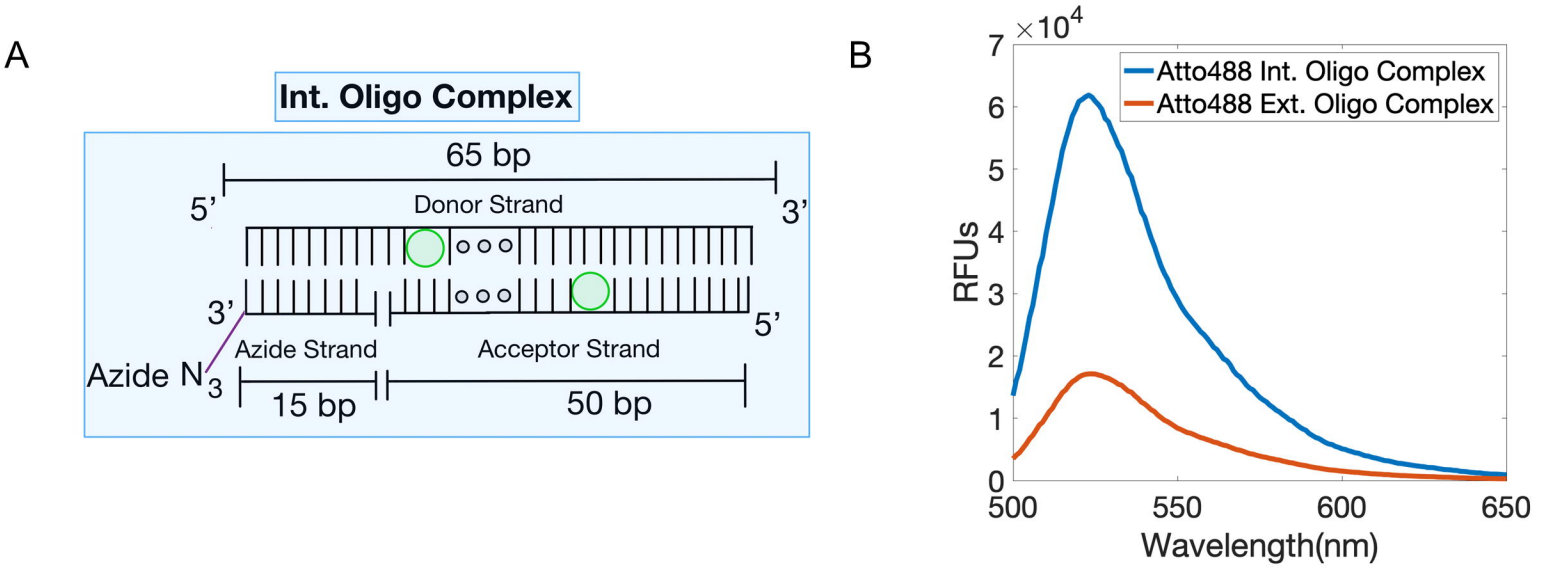


Figure 3. The int. labeling method increases intensity relative to the ext. labeling method. (A) Int. oligo complex containing the Int. Donor and Acceptor strands and the Azide strand. (B) Comparison of relative fluorescence intensity of the Atto488 probes using the int. and ext. oligo complexes (470 nm excitation). Data are representative from triplicates.

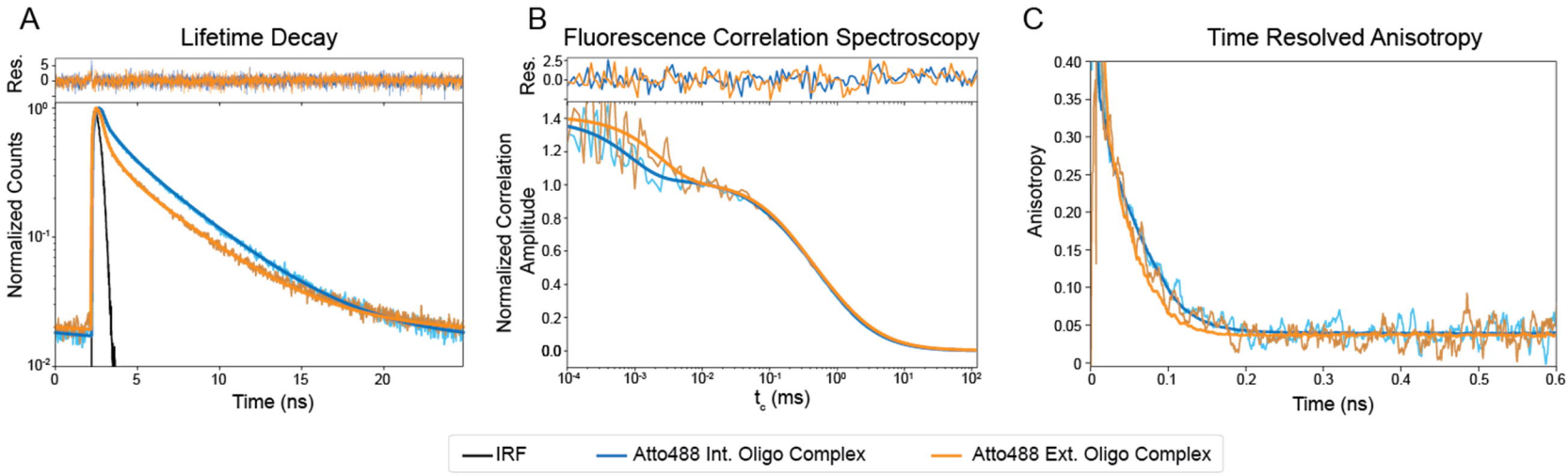


Figure 4. Differentiating static and dynamic quenching by Time-Resolved Fluorescence and Fluorescence Correlation Spectroscopy. Atto488 int. oligo complex fitting is shown in blue, with the raw data in light blue and Atto488 ext. oligo complex fitting is shown in orange, with the raw data in light orange. (A) Normalized fluorescence decays between Atto488 int. oligo complex and Atto488 ext. oligo complex. The difference in fluorescence lifetimes is visible by the difference in the slope of the decays. Residuals for the fitting model are shown on top. (B) Normalized fluorescence correlation between Atto488 int. oligo complex and Atto488 ext. oligo complex. The difference in dark triplet states is visible in the offset between the start of the curves to correlation times (t_c) ~10^{-3} ms. The curve overlaps between t_c being 10^{-1} and 10, indicating similar diffusion coefficients between samples. Residuals for the fitting model are shown on top. (C) Anisotropy rotational correlation between Atto488 int. oligo complex and Atto488 ext. oligo complex. The slight offset in rotational times could be due to differences in flexibility of the int. and ext. oligo complexes.

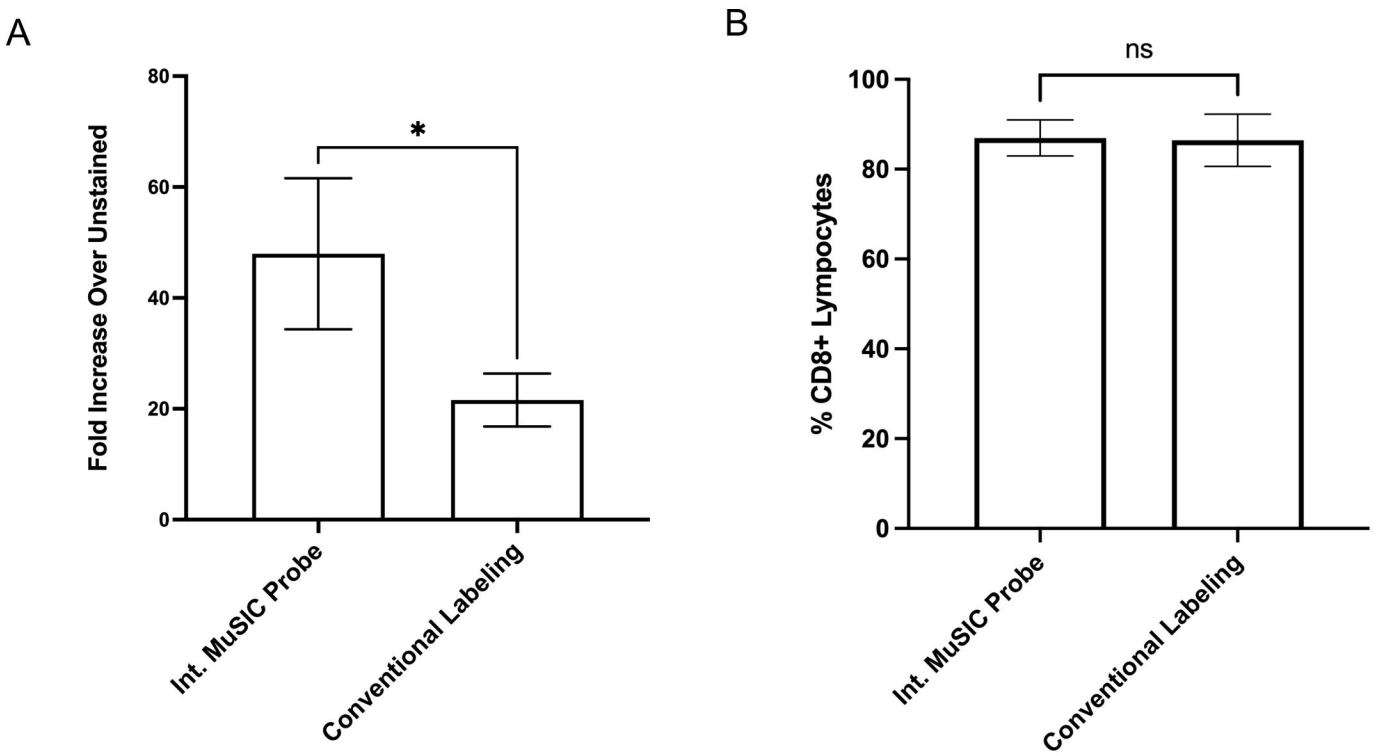


Figure 5. Comparing int. oligo complex and conventionally labeled antibodies in cell-based experiments. (A) Fold median signal increase of PBMCs stained with Atto488 int.-labeled CD8 antibodies or CF488A conventionally-labeled CD8 antibodies over unstained PBMCs. (B) Percentage of CD8+ lymphocytes in PBMC for int.-labeled CD8 antibodies compared to CF488A conventional-labeled CD8 antibodies. Measurements are in triplicate and error bars are standard error.

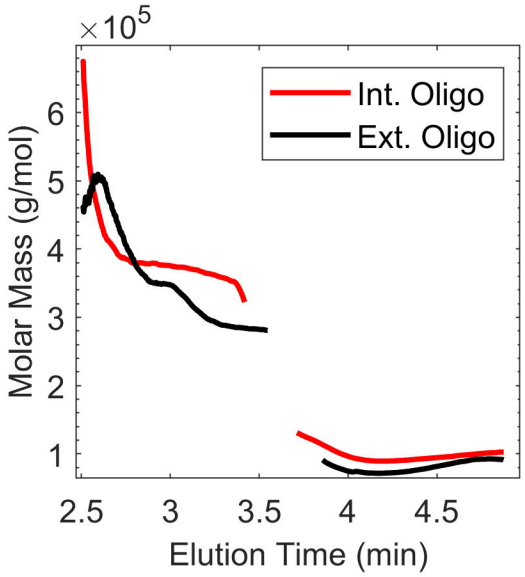


Figure S1. SEC-MALS Data for Labeled Antibody Purification. The molar mass versus elution time measured by SEC-MALS of Int. Oligo labeled antibody (red) and Ext. Oligo labeled antibody (black) solutions. Eluent from 2.5-3.5 minutes was collected for further analysis.

Appendix B

Biophysics Concept Inventory Survey: An Assessment in Biophysical Undergraduate Education

Biophysics Concept Inventory Survey: An Assessment in Biophysical Undergraduate Education

Danielle R. Latham[1], Joshua D. Alper[1,2,3,§], Hugo Sanabria[*]

[1] Department of Physics and Astronomy, Clemson University, Clemson, SC, 29634, USA
[2] Department of Biological Sciences, Clemson University, Clemson, SC, 29634, USA
[3] Eukaryotic Pathogen Innovations Center, Clemson University, Clemson, SC, 29634, USA

[*] Corresponding author: Hugo Sanabria (hsanabr@clemson.edu)
[§] Current address: GSK, 200 Cambridgepark Drive, Cambridge, MA, 02140

Abstract

Since the Force Concept Inventory in 1992, many concept inventories have been developed to cover classical scientific fields. However, there is a lack of concept inventories for interdisciplinary fields, such as biophysics. We introduce a Biophysical Conceptual Inventory Survey (BCIS), a 20-question, multiple-choice survey to measure student gains in biophysical concepts. The BCIS contains five question classifications: remember, understand, apply, analyze, and create as well as question concepts divided into primarily physics or primarily biology questions. We administered the BCIS to three cohorts of students over four years. Each cohort participated in a 10-week summer Research Experience for Undergraduates (REU) in biophysics. We compared the pre-survey (before REU) and post-survey (after REU) scores to determine the fraction of the maximum possible gain or loss realized. Our analysis of the results suggests that the BCIS shows no biases based on sex or ethnicity. Further, we used the BCIS to show that 69% of the REU participants showed gains in biophysics concepts, with majority of the total participant mean of gain occurring at the higher levels of Bloom's Taxonomy: create and analyze. Overall, participants obtain slightly higher scores in physics (8% increase) than biology (5% increase) when comparing the pre- and post-scores. The COVID-19 pandemic allows a splitting of pre-pandemic and post-pandemic cohorts, with the post-pandemic cohort showing significantly larger gains than the pre-pandemic students. These results show the BCIS, with question classifications and concepts, probes the students' ability to apply knowledge to various biophysical science topics without underlying biases and enables instructors to get answers to important questions about the effectiveness of their educational programs. The BCIS fills a gap for interdisciplinary concept inventories.

Introduction

Concept inventories exist for traditional fields in science (1-3), technology (4,5), engineering (6,7), and mathematics (8,9), but there remains a gap for interdisciplinary fields such as biophysics. Concept inventories arise from the need to have metrics to determine the depth of common student misunderstandings in sciences. Multiple choice questions allow instructors a quick, easy-to-grade method of probing their classes for complex topics. Several studies have already shown the benefits and logic of concept inventories (10) and how they can be best applied (11-14).

We developed a Biophysics Concept Inventory Survey (BCIS) to probe student learning across disciplines by generating 20 multiple-choice questions, which take an interdisciplinary approach to physics and biology. The BCIS contains five question classifications based on Bloom's Taxonomy: remember, understand, apply, analyze, and create (13,14). The question classification allows the probing of students' ability to apply biophysical concepts to various problems. Moreover, we classified the questions as either prominently probing physical or biological concepts. Results inform instructors in which concepts students struggle the most. We tested the BCIS for underlying biases using gathered demographic information, including sex and ethnicity of the REU participants.

I. Scientific and Pedagogical Background

Students have many incorrect ideas and misconceptions regarding science (15-17). Written exams and student interviews help determine these misconceptions, but they are long and take time to perform and analyze. In 1985, Halloun and Hestenes developed a multiple-choice concept inventory regarding the physics of motion to quickly determine student misconceptions (18). Questions for this survey were multiple choice, with one correct answer and several incorrect answers designed to distract. These distractor answers are designed from common misconceptions based on common student answers in written essays and student interviews (19). Shortly after the motion concept inventory, the force concept inventory was developed (20). The force concept inventory showed that students could recite Newton's third law but not apply it correctly. These early concept inventories led to an overturn of science education (10,12).

Since the first release of concept inventories, particularly the force concept inventory, there have been several studies showing the benefits and logic of concept inventories and how they can be best applied (11,20-22). Multiple choice questions allow administrators a quick, easy-to-grade means of probing student learning in complex topics. Concept inventories serve as a valuable tool for assessing the level of student comprehension and misconceptions in the field of sciences. (17,23,24).

Administrating concept inventories several times throughout a course allows instructors to determine student education progress during instruction. Typically, this change involves students doing better on the concept inventory after instruction showing an increase in score. The change in score gives a measure of how much information students gain after instruction. Often gain is the metric used to determine student advancement in a course. While there are ongoing discussions regarding the best way to calculate gain (25,26), gain is typically on a scale between zero and one, with a traditional, semester-long lecture course gives an average gain of ~0.25 (12).

Halloun and Hestenes work helped guide the creation of future concept inventories giving way to numerous concept inventories in multiple disciplines including physics (20,21,27,28), chemistry (29,30) and biology (1,2,31-34). However, these concept inventories are very specific, often covering a single topic within a single discipline such as kinematics (27) or electrostatics (28) from physics and natural selection (2) from biology. There are several concept inventories for traditional fields, but there remains a lack of tools for measuring student learning and understanding in interdisciplinary fields such as biophysics. We developed a Biophysics Concept Inventory Survey (BCIS) to address this need.

We developed the Biophysics Concept Inventory Survey (BCIS) to assess student understanding across disciplines by generating 20 multiple-choice questions that take an interdisciplinary approach to physics and biology. We wrote questions to be classified as primarily physics-based or primarily biology-based topics to inform instructors which topics cause students to struggle. Physics questions are typical physics concepts, including diffusion, kinetics, force and energy, density, pressure, mechanics,

electrostatics and optics, applied to a biological system, such as switching a walking person in a kinetics question to a cargo vesical moving along a microtubule. Biology questions put core biological concepts in the front, including molecular biology, genetics, and biochemistry, with less emphasis on the physical properties of biomolecules. As guidance, we modified some questions from previously existing concept inventories. For example, one question comes from the force concept inventory (20) where the original question involving the forces between charges spheres. Our modified version creates the situation as proteins are embedded in a cellular membrane. Biology based questions came from general biology concept inventories with questions being definition based and mechanism driven.

Additional design of the BCIS included considering Bloom's taxonomy of human cognition. We group each BCIS question into one of five classifications: remember, understand, apply, analyze, and create (13). These classifications enable instructors to probe students' ability to apply biophysical concepts at various cognition levels. It is not enough to repeat previous facts, but students should be able to use their knowledge to further their research and assist with troubleshooting. We want students to form problem-solving and logic skills. Addressing the questionnaire as a 'survey' helps students answer honestly and address test anxiety (35,36).

We calculated each participant's gain or loss of knowledge and then averaged the gains and losses together for an average of gains. We tested the BCIS for biases against sex and ethnicity. There were no significant differences between sex or ethnicity. Our study was interrupted by the COVID-19 pandemic. This interruption allowed us a unique opportunity to demonstrate how the BCIS distinguished between pre- and post-pandemic cohorts. Our results, for the pilot REU group, imply the BCIS can be used to determine the change in student understanding and application over time using multiple-choice questions for quick and easy grading. Thus, the BCIS fulfills a need for interdisciplinary evaluations across biophysics courses.

II. Methods and Statistical Tests

A. Biophysics Concept Inventory Survey (BCIS)

The Biophysics Concept Inventory Survey (BCIS) consists of 20 - multiple choice questions with a single correct answer. We used the force concept inventory (20) as example. For questions classified as primarily Physics, we use applications of physics concepts to biological systems. For example, instead of a charged particle, we used a charged DNA. Simple explanations were changed to have a biological context, such as changing a charged surface into a charged cellular membrane and making kinetic graphs related to movement within a cell or along a microtubule. Questions that are primarily Physics assess concepts on diffusion, kinetics, force and energy, density, pressure, mechanics, electrostatics, and optics. For questions classified as primarily Biology, we asked semiquantitative questions focused on Molecular Biology, Genetics and Biochemistry.

Instructor's access to the BCIS can be requested by filling out a google form with proof of instructors' role (37).

B. Research Experience for Undergraduates Sample Group

As a pilot test, we administered the BCIS to 32 students from three cohorts of undergraduate researchers who participated in the NSF funded REU Site: "Nature's machinery through the prism of physics, biology, chemistry and engineering" at Clemson University. The REU committee, consisting of the Primary Investigators of the REU site, and a faculty mentor, screened the applications to satisfy the programmatic goals of equal participation from participants with background in biological sciences and from the physical sciences. For each cohort, the REU committee balanced participation from URMs, sex, and from non-research-intensive institutions. Final assignment to the project was equally weighted the participant's interest and a final interview with the potential mentor. Recruitment was nationally, but with emphasis from the southeast. The participants came from 17 states, from private and public institutions of higher education ranging from primarily undergraduate institutions (PUI) to doctoral universities with very high research activity according to the Carnegie Classification. As part of their application, we gathered demographic information on the participants, such as sex and ethnicity.

The first week of the REU program, undergraduate researchers participated in a "Biophysics Bootcamp." During Bootcamp, participants participated in approximately 13 hours of traditional lectures and 17 hours of laboratory work, including introduction to research lectures. Participants spent this first Bootcamp week becoming familiar with Clemson University's campus, socializing with each other, and learning essential research and basic laboratory skills, such as how to keep a laboratory notebook and research safety, among other required introductions before they enter a laboratory setting. Additionally, each cohort received training in basic experimental and computational tools following a designed theme. For example, in 2021, participants determined the 'size' of GFP by various means, including Fluorescent Correlation Spectroscopy, Size Exclusion Chromatography, Computational simulations using Visual Molecular Dynamics (38), and quantitative analysis of SDS-PAGE gels. After the Bootcamp participants wrote a report formatted as a Biophysical Journal article. This training helped participants understand experimental validation through many means and determine the differences (pros and cons) of different experimental designs.

For the remaining 9 weeks of the REU program, participants worked on collaborative, interdisciplinary research projects in pairs, but with individual and unique project objectives, where one undergraduate researcher had an experimental focus, while the other had a computational aspect of the same problem; or one undergraduate researcher was in a physics laboratory and the other was doing the more biological aspects of the project. This approach allowed participants to build collaboration skills while gaining exposure to both experimental and computational approaches to research.

To supplement their experience and aid in building their professional development skills (39,40), REU participants had weekly meetings with their cohorts at which they presented research updates including project design, background, and importance. Participants also met weekly for a journal club at which they took turns presenting recently published research articles relating to their project to encourage staying up-to-date on relevant research for their research topic and practicing critical reading of the literature. There were also weekly professional seminars given by experts at the university covering topics such as scientific writing, networking, and conflict resolution. At the end of the summer, undergraduate researchers participated in Clemson University's Undergraduate Research Symposium.

During the REU, participants did not have any traditional classroom instruction regarding the topics covered by the BCIS and participants were not quizzed or given traditional homework, such as problem sets. The BCIS was developed separately from the REU curriculum. Participants drove their learning by finding and reading the scientific literature, asking questions of those around them, and problem-solving on their research projects. Thus, this sampling is biased towards undergraduate researchers who participated in an interactive, experiential learning approach (22,41), instead of students who participated in a traditional, semester-long lecture course.

C. Administration of the survey

Participants took the BCIS upon arrival (pre-survey) to the REU site and upon departure (post-survey). The question order remained the same for the pre-survey and post-survey to ensure the order of the question played no part in answer changes between the pre- and post-surveys. Access to the survey required a password and Respondus LockDown Browser to ensure the survey was given to all participants simultaneously with no outside resources. Participants had 35 minutes to answer the 20 questions.

D. Matched Data

We used matched data (42) for all analyses, allowing the consideration of participant demographics. Therefore, participant data calculations are completed for each individual and then pooled via demographics to form statistical groups.

E. Fraction of Maximum Possible Gain Realized

For each participant, we calculated the pre- and post-scores from the pre- and post-survey, respectively. Each question was weighted the same with typical grading procedures to determine the score;

the number of correctly answered questions was divided by the total number of questions to give a percentage answered correctly.

Eq. (1) shows how we compared the pre- and post-scores for each participant to obtain a Gain, No Change, Loss (GNL) value. We calculated the Fraction of the Maximum Possible Gain Realized (gain) (12) for participants who scored higher on the post-survey than the pre-survey. For participants who scored lower on the post-survey than the pre-survey, we calculated the Maximum Possible Loss Forfeited (loss). While the concept of loss has been deliberated before (25,26), our loss calculation method is normalized regarding the percent of questions answered incorrectly compared to what was initially known. The participant is assigned a zero when the pre- and post-scores are identical, signaling no change. There were no participants who obtained a perfect score (100%) on the pre- or post-survey. Therefore, their gains, no change and loss are calculated as follows:

$$\begin{aligned} \text{Gain} &= \frac{\text{Post-Score} - \text{Pre-Score}}{100\% - \text{Pre-Score}} \quad , \quad \text{Post-Score} > \text{Pre-Score} \quad (a) \\ \text{No Change} &= 0 \quad , \quad \text{Post-Score} = \text{Pre-Score} \quad (b) \\ \text{Loss} &= \frac{\text{Post-Score} - \text{Pre-Score}}{\text{Pre-Score}} \quad , \quad \text{Post-Score} < \text{Pre-Score} \quad (c) \end{aligned} \tag{1}$$

With a mean of GNL (Gain or G, No change or N, Loss or L) that is the weighted average of the three possible scores as

$$\langle GNL \rangle = \frac{n_G}{n_T}\left(\frac{1}{n_G}\sum_{n_G} G\right) + \frac{n_N}{n_T}\left(\frac{1}{n_N}\sum_{n_N} N\right) + \frac{n_L}{n_T}\left(\frac{1}{n_L}\sum_{n_L} L\right) = \frac{1}{n_T}\left(\sum_{n_G} G + \sum_{n_N} N + \frac{1}{n_L}\sum_{n_L} L\right) \tag{2}$$

The mean GNL method creates a scale from -1 (total loss) to 1 (total gain), where negative numbers represent loss and positive numbers represent gain. This method assists in averaging statistics and further data analysis.

F. P-Values and Effect Size

Participants were deidentified and grouped into different demographic groups: sex, underrepresented minority (URM) status, and college major as were self-reported by the participants. We calculated and considered Cohen's effect size (*d*, Eq. 3) (43-45) to compare between groups. The effect size shows the size of the shift between the pre- and post-scores. We opted for Cohen's effect size since it provides a good measure for smaller sampling sizes, and we have a total sample size of 32. The effect size is calculated by Eq. (3) where "PooledSTD" is the pooled standard deviation of all the pre- and post-survey scores.

$$\text{Effect Size} = \text{d} = \frac{\langle \text{Post-Score} \rangle - \langle \text{Pre-Score} \rangle}{\text{PooledSTD}}, \tag{3}$$

where the brackets "⟨ ⟩" represent the mean. In this manner, an effect size of 0.2 is a small shift, 0.5 is a medium shift, and 0.8 is a large shift (43,46).

Further, each demographic grouping was compared using Student's t-test. For each t-test, we used normal quantile-quantile plots to ensure the sampling data distribution was close to normal. With such a small sampling size, a p-value may not be efficient for determining the differences between subgroups (47), but a combination of p-values and effect size allows a complete comparison between various subgroups for this study (48). We considered p-value < 0.10 to be statistically significant.

G. Question Subject Percentages

To determine the effect size regarding subject matter, questions of similar subjects were grouped together. Then the total mean and standard deviation for the pre- and post- responses were determined for

each group. This allowed the pooled standard deviation and effect size for each grouping to be determined.

III. Results and Discussion

A. The BCIS shows medium gains from REU participants

Overall, the average BCIS scores increased by 7% from the pre-score (49.4% ± 14.2%) to the post-score (56.4 ± 13.1%). With a pre-score of 50% (the mean pre-score is not statistically different from 50%, p-value = 0.8), the BCIS is easy enough for undergraduate students to feel confident while leaving enough room for students to achieve gain.

We found that the 32 REU participants had a mean of gain, no change, loss (GNL) of 0.13 ± 0.18 and an effect size of 0.51 with three groups: Gain with 22 participants, Loss with 6 participants, and No Change with 4 participants (**Figure 1**). Gain participants have a large effect size of 0.98 with a mean of gains 0.23 ± 0.11 (n = 22). Loss participants have a medium effect size of -0.40 with a mean of loss -0.15 ± 0.07 (n = 6).

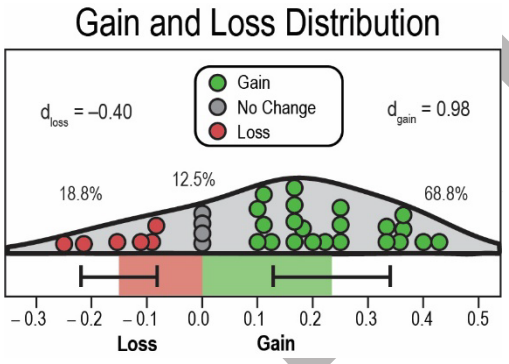


Figure 1 Overall Gain and Losses. The distribution of the 32 REU students' gain or loss values shown in the horizontal violin chart where each point represents a student's GNL score (*green* represents gain, *gray* represents no change, and *red* represents loss). The percentage and effect size (*d*) for each group can be found above the respective group. The horizontal bar chart below each group shows the mean with error bars representing the standard deviation for gain and loss groups.

The increase in gain and effect size may be attributed to interactive experimental learning and may not reflect a traditional lecture course (49,50). Many previous studies discard the students with losses(51). Here we decided to divide the gains and losses but show both groups (26). The 69% of participants benefited from the REU as assessed by the BCI with overall positive gains.

B. The BCIS can identify students' weak and strong subjects

We analyzed the BCIS results by question subject. While the questions are interdisciplinary, we classified each question as a principal biology subject (6 questions) or a physics subject (14 questions). Further, the questions address specific topics, including Kinetics, Mechanics, Force/Energy, Electrostatics, Density, Pressure, Diffusion, Optics for physics and Molecular Biology, Genetics, Biochemistry for biology. We compared the pre-survey and post-survey responses for each student to identify the topics that individual participants either better understand or continue to struggle with after instruction (Figure 2).

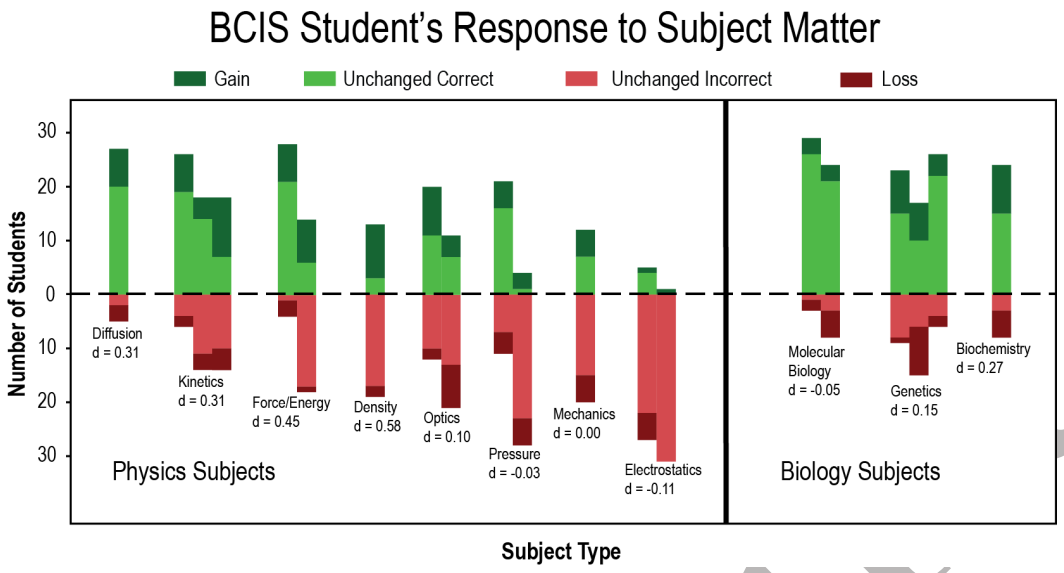


Figure 2 BCIS contains a variety of question categories that probe a breadth of physics and biology subjects. Bar graphs represent student gain or loss for each question, with labels indicating the main **subject** the question is covering. Students in *dark green* are marked as 'Gain', answering correctly only on the post-survey. Students in *light green* are marked as 'Unchanged Correct', answering correctly on the pre- and post-survey. Students in *light red* are marked as 'Unchanged Incorrect' answered incorrectly on both the pre- and post-surveys. Students in *dark red* are marked as 'Loss', answering correctly on the pre-survey, but incorrectly on the post-survey.

We found that the participants came in to the REU program already understanding biology subjects better than physics subjects, with 69% of answers correct for biology subject questions on the pre-survey compared to only 41% for physics subjects. Participants had a slightly larger effect size for physics (d = 0.16) compared to biology (d = 0.12), but both show small shifts. A closer look showed participants shifted more on certain subjects than others. Within the physics group of questions, we found participants showed larger shifts in introductory physics concepts, like density (d = 0.58), forces/energy (d = 0.45), and kinetics (d = 0.31), with small negative shifts, denoting losses, in more advanced physics concepts, like electrostatics (d = -0.11), and pressure (d = -0.03). The small losses may be attributed to guessing, due to the nature of multiple-choice testing (52). We observe smaller changes regarding biology subjects. Biochemistry (d = 0.27) showed the largest total change, with molecular biology (d = -0.05) showed a slight negative shift.

C. The BCIS uses question classifications to assess participants' understanding

Typical assessments only tend to probe student knowledge – their ability to repeat back information previously given. It is crucial to ensure that students can apply this knowledge. Thus, we designed the BCIS with questions across multiple levels of Bloom's Taxonomy of Educational Objectives (13,14) (**Table 1**).

Table 1 Bloom's Taxonomy: BCIS Question Classification Details

Classification	Number of Questions	Description of Question Classification
Remember	7	Recognizing and Recalling Information
Understand	3	Interpreting, Explaining, Summarizing
Apply	3	Applying rules, methods, or principles to new situations

Analyze	3	Classifying and understanding components parts within a system
Create	4	Creating new connections and combining ideas
Evaluate	0	Addressing controversies, forming opinions

We found that the REU students showed nearly zero mean of GNL at the lower levels of Bloom's taxonomy: remember (0.06 ± 0.37), understand (0.06 ± 0.45) and apply (0.05 ± 0.43). However, we found considerable gains in create (0.22 ± 0.39) and analyze (0.23 ± 0.50) (**Figure 3**). We attribute these results to the active learning approach of the REU. The experience increased participants' ability to apply knowledge, particularly regarding creating new connections (create) and understanding how system parts fit together (analyze). However, without required reading, traditional problem sets, or classroom-based lectures, participant baseline remember showed no gain.

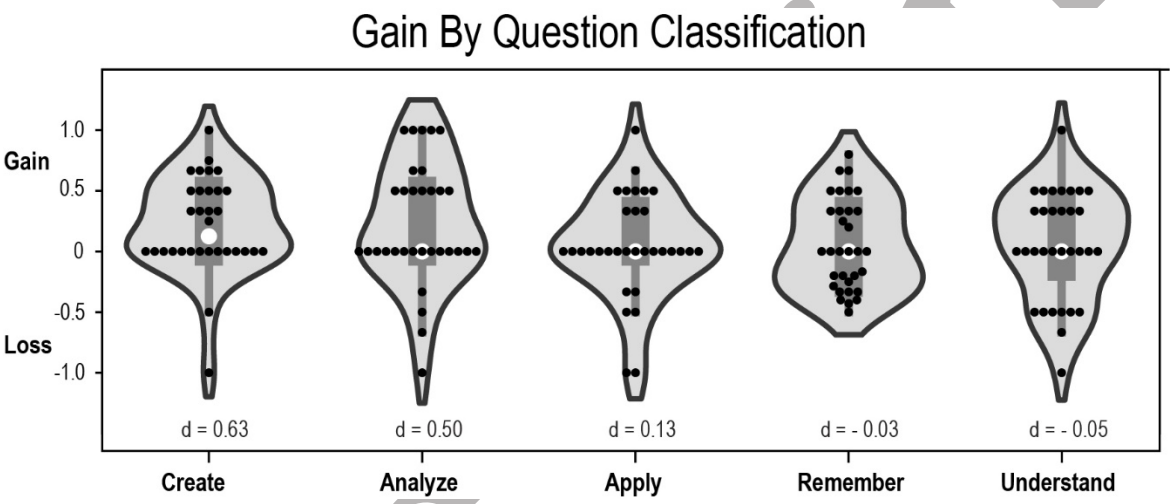


Figure 3 The BCIS contains a variety of question categories that probe different levels of understanding. Violin plots show student gain for each question category. Each *black* dot represents a student. The box-whiskers plots inside the violin plots show the quartiles, with the *white* dot representing the median GNL. Effect size (*d*) is shown below the plots.

D. This BCIS is non-bias for Sex and URM Status but shows a preference for College Major

We pooled all REU cohorts by demographic information to test the BCIS for biases (53,54). We found no statistically significant differences in gain or loss corresponding to the participants' sex (two-tailed t-test, p-value = 0.90) or Under Represented Minority (URM) status (two-tailed t-test, p-value = 0.62). The effect size for sex was 0.41 for males and 0.61 for females, indicating that both groups showed medium gains from the REU (**Figure 4A**). The effect size for URM status was 0.67 for URM participants and 0.45 for non-URM participants, indicating that both groups also showed medium gains from the REU (**Figure 4B**). Additionally, this data supports the conclusion that the BCIS is not inherently biased based on gender or ethnicity.

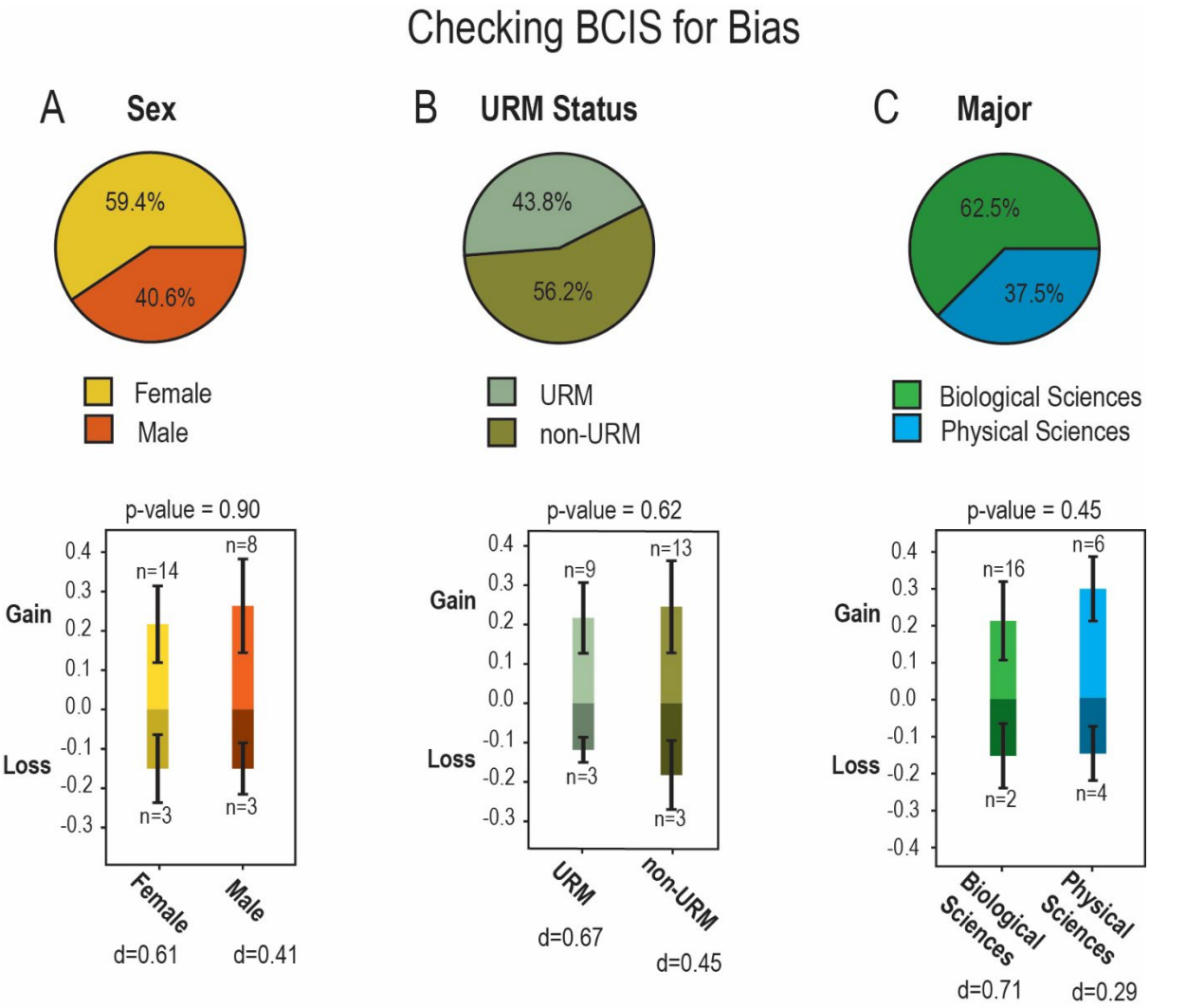


Figure 4: The BCIS shows no bias for sex or URM. Pie charts show the distribution of student demographics based on (A) sex, (B) Under Represented Minority (URM) Status and (C) college major. Bar charts show the mean of the gain (positive) and loss (negative) with the error bars denoting standard deviation for each distribution. The effect size (*d*) is shown below for each demographic. Sample size for each group is denoted by 'n =' near the error bar for that group. The four students with no change are not shown in the bar charts.

Further, we pooled the participants into their respective college majors: physical sciences (for participants majoring in physical sciences or engineering) or biological sciences (for participants majoring in any of the life sciences). Approximately two-thirds of participants have a biological sciences undergraduate major (**Figure 4C**). We found no statistically significant differences in Gain or Loss corresponding to the participant's major (two-tailed t-test, p-value = 0.45).

However, the effect size is 0.71 for biological sciences showing large growth, while physical sciences only have an effect size of 0.29 showing small growth, which implies that biological science majors experienced bigger gains than physical science majors during the REU. This could be due to biological sciences lacking more physics knowledge than physical sciences lacking biology knowledge at the start of the program.

E. Example: COVID-19 Impact on Participant Gains

We first administered the BCIS to REU participants in 2019; however, a worldwide pandemic interrupted and altered the study's course as safety concerns postponed the 2020 REU. We offered deferment to those participants we had accepted to the 2020 REU. Thus the 2021 REU cohort consisted of a mix of participants who were accepted pre-pandemic (2020; n = 7) and post-pandemic (2021; n = 8). Our analysis of survey data and conversations with participants revealed that the participants who had applied post-pandemic (in 2021 and 2022) lacked traditional laboratory courses that would have accompanied the introductory science courses at their home institutes, while those who had applied pre-pandemic (in 2019 and 2020) had those lab courses. This distinction led us to divide the participant cohorts into "pre-pandemic" and "post-pandemic" groups.

We found that the pre-pandemic group had a mean of GNL 0.07 ± 0.18 with an effect size of 0.35 (n = 14), and the post-pandemic group had a mean of GNL 0.18 ± 0.18 with an effect size of 0.69 (n = 18). A comparison between the two groups showed they were significantly different (2-tailed t-test, p-value = 0.09). These differences are explained by both larger gains (Fig 5A) and a greater fraction of students showing gain (Fig 5B) in the post-pandemic group.

A more detailed inspection of this data using the question classifications (**Figure 5C**) shows similar, medium effect sizes indicating gain for both pre- and post-pandemic cohorts at the higher levels of Bloom's taxonomy: Create and Analyze. However, there are significant differences at the lower levels of Bloom's taxonomy, with the pre-pandemic group showing negative effect sizes in Understand and Remember, implying a loss. In contrast, the post-pandemic group shows small, positive shifts in these classifications.

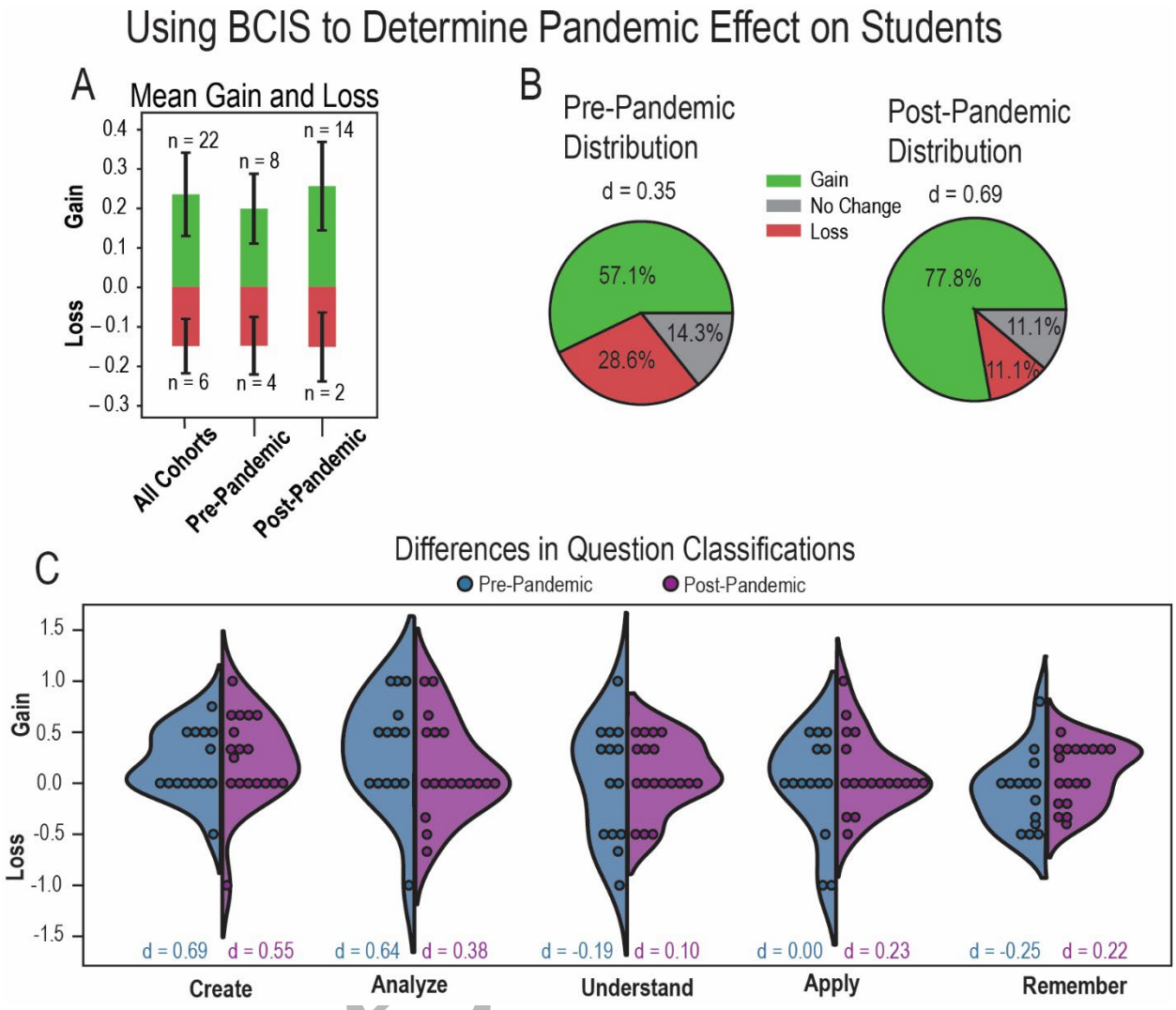


Figure 2 Applying the BCIS before and after the COVID-19 global pandemic suggest an increased benefit of experiential learning opportunities for post-pandemic REU students. (A) Bar graph shows mean gain and loss for the pre- and post-pandemic cohorts, where the error bars are standard deviation. (B) Pie charts show fractions of the pre- and post-pandemic cohorts with gain and loss. The four students with no change are not shown in the bar charts. (C) In the violin plots, each student is represented by a point. The left side (*blue*) of each violin plot represents the pre-pandemic cohorts, while the right side (*violet*) of each violin plot represents the post-pandemic cohorts. The effect size (*d*) under each corresponding distribution shows the differences in the pre and post BCIS responses for each distribution.

Together, these results show a distinction between pre-pandemic and post-pandemic cohorts, including a 21% increase in the number of participants who exhibited gain post-pandemic, larger effect sizes for questions classified lower Bloom's taxonomy (understand, remember, apply) for post-pandemic participants, and an overall 0.34 increase in effect size and 0.11 increase in gains for post-pandemic compared to pre-pandemic. These results imply educational disruption has interfered with student education, but hands-on, active learning approaches, like summer REU experiential learning programs, may aid in recovery. They suggest that immersive lab experience benefits students, with the exposure helping return students to a better, pre-pandemic learning state.

IV. Conclusion

The current concept inventories are lacking for interdisciplinary fields. To fill this gap, we created the Biophysics Concept Inventory Survey (BCIS). We administered the BCIS to 32 REU participants as a pilot group. By having different question classifications and subject material, we could better understand participants' weak points, including second semester physics topics such as electrostatics, pressure and optics as well as applied biological subjects such as molecular biology. Additionally, the BCIS results suggest that experiential learning through an REU leads to higher means of GNL for at the higher end of Bloom's Taxonomy (create and analyze) than at the lower end (remember and understand). Applying the BCIS to traditional lecture courses would be interesting, as we anticipate larger gains at the lower end of Bloom's Taxonomy. For traditional semester-long (~16 weeks) courses, it is likely best to apply the BCIS three times: at the start of the semester, halfway through the semester, and at the end of the semester (55). The classification of questions by subject and Bloom's taxonomy level allows instructor to determine what students are struggling with and adjust course direction at a midway point of a course.

This study suggests that the surveys like the BCIS are useful tools to evaluate student gains in interdisciplinary courses and active learning experiences. However, our results are limited to a small number of REU participants. Therefore, we must administer it to more students for a larger sample size. After the BCIS is robustly tested on a larger sample size many potentials open up, such as (a) building a database of questions, (b) probing class progression at a midpoint (c) checking student's pervious understanding of physics and biology with topic specific concept inventories. Instructors who want to apply this to their course or research program can request by contacting the authors of this study and filling out a google form with a proof of instructor's role.

In conclusion, the BCIS starts to fill the need for an interdisciplinary method of evaluating student progress in biophysics courses. It is unbiased in measuring interdisciplinary biology and physics understanding. It covers various subjects in physics and biology, allowing the understanding of students' weak points. Question classifications based on Bloom's Taxonomy grant the ability to understand students' level of knowledge and the ability to apply that knowledge. In our pilot study, we found apparent differences in performance on the BCIS between pre-pandemic and post-pandemic REU undergraduate researchers. In the future, we will expand the BCIS by adding an extensive databank of questions, enabling instructors to customize the balancing of the BCIS by classification, subject matter and question type. A databank would allow instructors to build a specialized concept inventory for their class, covering topics they find could be more relevant to specific needs.

V. IRB Statement

This work, which is classified as exempt under category 2 in accordance with federal regulations 45 CFR 46.104(d), was carried out in accordance with the standards established by Clemson University. 2018270

VI. Author Contributions

DL wrote the manuscript with revision by JA and HS. DL did the data analysis. JA and HS supervised the work and provided guidance to DL. JA and HS designed the BCIS. HS lead the administration of the BCIS to the REU participants. JA and HS are directors of the REU Site "Nature's machinery through the prism of physics, biology, chemistry and engineering" at Clemson.

VII. Acknowledgments

We thank Celeste Hackett for helping on the organization of the NSF funded REU Site: "Nature's machinery through the prism of physics, biology, chemistry and engineering". Special thanks to the Bootcamp teaching assistances of all cohorts for leading and modifying protocols to match the students' needs: Zhenzhen Zhang, George Hamilton, Katherine Wentworth, Rajen Goutam, Subash Godar, Narendar

Kolimi, Ashok Pabbathi, and Vincent Clanzy III. A large thanks to all the Primary Investigators and Guest Lectures who mentored the REU participants through their summer at Clemson.

This research was supported by the National Science Foundation (NSF) award DBI-1757658 to Drs. Alper and Sanabria. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the NSF. The funders had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; and in the preparation, review, or approval of the manuscript.

References

1. Smith, M. K., W. B. Wood, and J. K. Knight. 2008. The Genetics Concept Assessment: A New Concept Inventory for Gauging Student Understanding of Genetics. *CBE—Life Sciences Education*. 7(4):422-430, doi: 10.1187/cbe.08-08-0045, https://www.lifescied.org/doi/abs/10.1187/cbe.08-08-0045.
2. Anderson, D. L., K. M. Fisher, and G. J. Norman. 2002. Development and evaluation of the conceptual inventory of natural selection. *Journal of research in science teaching*. 39(10):952-978.
3. Evans, D. L., G. L. Gray, S. Krause, J. Martin, C. Midkiff, B. M. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif, R. Streveler, and K. Wage (2003). Progress on concept inventory assessment tools. 33rd Annual Frontiers in Education, 2003. FIE 2003.
4. Caceffo, R., S. Wolfman, K. S. Booth, and R. Azevedo (2016). Developing a Computer Science Concept Inventory for Introductory Programming. Proceedings of the 47th ACM Technical Symposium on Computing Science Education. Association for Computing Machinery.
5. Almstrum, V. L., P. B. Henderson, V. Harvey, C. Heeren, W. Marion, C. Riedesel, L.-K. Soh, and A. E. Tew (2006). Concept inventories in computer science for the topic discrete mathematics. Working group reports on ITiCSE on Innovation and technology in computer science education. Association for Computing Machinery.
6. Krause, S., J. C. Decker, and R. Griffin (2003). Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses. 33rd Annual Frontiers in Education, 2003. FIE 2003.
7. Prince, M., M. Vigeant, and K. Nottis. 2012. Development of the Heat and Energy Concept Inventory: Preliminary Results on the Prevalence and Persistence of Engineering Students' Misconceptions. *Journal of Engineering Education*. 101(3):412-438, doi: https://doi.org/10.1002/j.2168-9830.2012.tb00056.x, https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2012.tb00056.x.
8. Steif, P. S., and J. A. Dantzler. 2005. A Statics Concept Inventory: Development and Psychometric Analysis. *Journal of Engineering Education*. 94(4):363-371, doi: https://doi.org/10.1002/j.2168-9830.2005.tb00864.x, https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2005.tb00864.x.
9. Gleason, J., S. Bagley, M. Thomas, L. Rice, and D. White. 2019. The calculus concept inventory: a psychometric analysis and implications for use. *International Journal of Mathematical Education in Science and Technology*. 50(6):825-838, doi: 10.1080/0020739X.2018.1538466, https://doi.org/10.1080/0020739X.2018.1538466.
10. Treagust, D. F. 1988. Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*. 10(2):159-169, doi: 10.1080/0950069880100204, https://doi.org/10.1080/0950069880100204.
11. Huffman, D., and P. Heller. 1995. What Does the Force Concept Inventory Actually Measure? *The Physics Teacher*. 33(3):138 - 143, doi: 10.1119/1.2344171.

12. Hake, R. 1998. Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics - AMER J PHYS*. 66, doi: 10.1119/1.18809.
13. Bloom, B. S. 1956. Taxonomy of educational objectives: The classification of educational goals, 1st ed. Longman Group, Harlow, Essex, England.
14. Anderson, L. W., D. R. Krathwohl, and B. S. Bloom. 2001. A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives / editors, Lorin W. Anderson, David Krathwohl ; contributors, Peter W. Airasian ... [et al.]. Longman, New York.
15. Soeharto, S., B. Csapó, E. Sarimanah, F. I. Dewi, and T. Sabri. 2019. A Review of Students' Common Misconceptions in Science and Their Diagnostic Assessment Tools. *Jurnal Pendidikan IPA Indonesia*.
16. Driver, R. 1981. Pupils' Alternative Frameworks in Science. *European Journal of Science Education*. 3(1):93-101, doi: 10.1080/0140528810030109, https://doi.org/10.1080/0140528810030109.
17. Linke, R. D., and M. I. Venz. 1978. Misconceptions in physical science among non-science background students. *Research in Science Education*. 8(1):183-193, doi: 10.1007/BF02558690, https://doi.org/10.1007/BF02558690.
18. Halloun, I., and D. Hestenes. 1985. Common Sense Concepts About Motion. *American Journal of Physics - AMER J PHYS*. 53:1056-1065, doi: 10.1119/1.14031.
19. Tamir, P. 1971. An Alternative Approach to The Construction of Multiple Choice Test Items. *Journal of Biological Education*. 5(6):305-307, doi: 10.1080/00219266.1971.9653728, https://doi.org/10.1080/00219266.1971.9653728.
20. Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force Concept Inventory. *The Physics Teacher*. 30(3):141 - 158, doi: 10.1119/1.2343497.
21. Thornton, R. K., and D. R. Sokoloff. 1998. Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*. 66(4):338 - 352, doi: 10.1119/1.18863.
22. Korff, J. V., B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell. 2016. Secondary analysis of teaching methods in introductory physics: A 50 k-student study. *American Journal of Physics*. 84(12):969-974, doi: 10.1119/1.4964354, https://aapt.scitation.org/doi/abs/10.1119/1.4964354.
23. Fensham, P. J., J. Garrard, and L. West. 1981. The use of cognitive mapping in teaching and learning strategies. *Research in Science Education*. 11(1):121-129, doi: 10.1007/BF02356775, https://doi.org/10.1007/BF02356775.
24. Halloun, I. A., and D. Hestenes. 1985. The initial knowledge state of college physics students. *American Journal of Physics*. 53(11):1043-1055, doi: 10.1119/1.14030, https://doi.org/10.1119/1.14030.
25. Marx, J. D. 2007. Normalized change. *American Journal of Physics*. 75(1):87-91, doi: 10.1119/1.2372468.
26. Miller, K., N. Lasry, O. Reshef, J. Dowd, I. Araujo, and E. Mazur (2010). Losing it: The Influence of Losses on Individuals' Normalized Gains. Physics Education Research Conference 2010.
27. Beichner, R. 1994. Testing student interpretation of kinematics graphs. *American Journal of Physics*. 62(3):750 - 762, doi: 10.1119/1.17449.
28. Ding, L., R. Chabay, B. Sherwood, and R. Beichner. 2006. Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research*. 2(1):010105, doi: 10.1103/PhysRevSTPER.2.010105, https://link.aps.org/doi/10.1103/PhysRevSTPER.2.010105.
29. Landis, C., G. Lisenky, J. Lorenz, K. Meeker, C. Wamser, and A. Ellis. 2000. Chemistry Conceptests: A Pathway to Interactive Classrooms. Prentice Hall, Upper Saddle River (NJ).
30. Mulford, D. R., and W. R. Robinson. 2002. An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *Journal of Chemical Education*. 79:739-744.

31. Klymkowsky, M. W., and K. Garvin-Doxas. 2008. Recognizing student misconceptions through Ed's Tools and the Biology Concept Inventory. *PLoS biology*. 6(1):e3.
32. Williams, K. S., K. M. Fisher, D. L. Anderson, M. U. Smith, and J. E. Lineback (2008). Using diagnostic test items to assess conceptual understanding of basic biology ideas: A plan for programmatic assessment.
33. Odom, A. L., and L. H. Barrow. 1995. Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *Journal of research in Science Teaching*. 32(1):45-61.
34. Nehm, R. H., and L. Reilly. 2007. Biology majors' knowledge and misconceptions of natural selection. *BioScience*. 57(3):263-272.
35. Tryon, G. S. 1980. The Measurement and Treatment of Test Anxiety. *Review of Educational Research*. 50(2):343-372, doi: 10.3102/00346543050002343, https://journals.sagepub.com/doi/abs/10.3102/00346543050002343.
36. Hembree, R. 1988. Correlates, Causes, Effects, and Treatment of Test Anxiety. *Review of Educational Research*. 58(1):47-77, doi: 10.3102/00346543058001047, https://journals.sagepub.com/doi/abs/10.3102/00346543058001047.
37. Latham, D., J. Alper, and H. Sanabria. 2023. https://forms.gle/xsoHAQPyZe7XVFtm6
38. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 14(1):33-38, doi: https://doi.org/10.1016/0263-7855(96)00018-5, https://www.sciencedirect.com/science/article/pii/0263785596000185.
39. Linn, M. C., E. Palmer, A. Baranger, E. Gerard, and E. Stone. 2015. Undergraduate research experiences: Impacts and opportunities. *Science*. 347(6222):1261757, doi: doi:10.1126/science.1261757, https://www.science.org/doi/abs/10.1126/science.1261757.
40. Lopatto, D., S. Tobias, C. o. U. Research, and R. C. f. S. Advancement. 2010. Science in Solution: The Impact of Undergraduate Research on Student Learning. Council on Undergraduate Research.
41. Coletta, V. P., and J. A. Phillips. 2005. Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*. 73(12):1172-1182, doi: 10.1119/1.2117109, https://aapt.scitation.org/doi/abs/10.1119/1.2117109.
42. Cochran, W. G. 1968. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 24:295-313, doi: 10.2307/2528036.
43. Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Inc.
44. Cohen, J. 1992. A power primer. *Psychological Bulletin*. 112:155-159, doi: 10.1037/0033-2909.112.1.155.
45. Kelley, K., and K. J. Preacher. 2012. On effect size. *Psychological Methods*. 17:137-152, doi: 10.1037/a0028086.
46. Sawilowsky, S. S. 2009. New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*. 8:26.
47. Wasserstein, R. L., and N. A. Lazar. 2016. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 70(2):129-133, doi: 10.1080/00031305.2016.1154108, https://doi.org/10.1080/00031305.2016.1154108.
48. Nissen, J. M., R. M. Talbot, A. Nasim Thompson, and B. Van Dusen. 2018. Comparison of normalized gain and Cohen's d for analyzing gains on concept inventories. *Physical Review Physics Education Research*. 14(1):010115, doi: 10.1103/PhysRevPhysEducRes.14.010115, https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.14.010115.
49. Lestari, I. F. 2021. Experiential learning using STEM approach in improving students' problem solving ability. *Journal of Physics: Conference Series*. 1806(1):012005, doi: 10.1088/1742-6596/1806/1/012005, https://dx.doi.org/10.1088/1742-6596/1806/1/012005.
50. Khairati, K., W. Artika, M. A. Sarong, A. Abdullah, and H. Hasanuddin. 2021. Implementation of STEM-Based Experiential Learning to Improve Critical Thinking Skills on Ecosystem Materials.

Jurnal Penelitian Pendidikan IPA. 7(4):752-757, doi: 10.29303/jppipa.v7i4.850, https://jppipa.unram.ac.id/index.php/jppipa/article/view/850.

51. Pérez García, M., C. J. Luxford, T. L. Windus, and T. Holme. 2016. A Quantum Chemistry Concept Inventory for Physical Chemistry Classes. *Journal of Chemical Education*. 93(4):605-612, doi: 10.1021/acs.jchemed.5b00781, https://doi.org/10.1021/acs.jchemed.5b00781.
52. Andrich, D., I. Marais, and S. Humphry. 2012. Using a Theorem by Andersen and the Dichotomous Rasch Model to Assess the Presence of Random Guessing in Multiple Choice Items. *Journal of Educational and Behavioral Statistics*. 37(3):417-442, http://www.jstor.org/stable/23256848.
53. Oberai, H., and I. M. Anand. 2018. Unconscious Bias: Thinking Without Thinking. *Human Resource Management International Digest*. 26(6), doi: 10.1108/HRMID-05-2018-0102.
54. Dee, T., and S. Gershenson (2017). Unconscious Bias in the Classroom: Evidence and Opportunities. Stanford Center for Education Policy Analysis.
55. Lee, U. J., G. C. Sbeglia, M. Ha, S. J. Finch, and R. H. Nehm. 2015. Clicker Score Trajectories and Concept Inventory Scores as Predictors for Early Warning Systems for Large STEM Classes. *Journal of Science Education and Technology*. 24(6):848-860, doi: 10.1007/s10956-015-9568-2, https://doi.org/10.1007/s10956-015-9568-2.

Appendix C

Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies

Exequiel Medina[1,2], Danielle R. Latham[2] and Hugo Sanabria[2]

Evidence regarding protein structure and function manifest the imperative role that dynamics play in proteins, underlining reconsideration of the unanimated sequence-to-structure-to-function paradigm. Structural dynamics portray a heterogeneous energy landscape described by conformational ensembles where each structural representation can be responsible for unique functions or enable macromolecular assemblies. Using the human p27/Cdk2/Cyclin A ternary complex as an example, we highlight the vital role of intramolecular and intermolecular dynamics for target recognition, binding, and inhibition as a critical modulator of cell division. Rapidly sampling configurations is critical for the population of different conformational ensembles encoding functional roles. To garner this knowledge, we present how the integration of (sub)ensemble and single-molecule fluorescence spectroscopy with molecular dynamic simulations can characterize structural dynamics linking the heterogeneous ensembles to function. The incorporation of dynamics into the sequence-to-structure-to-function paradigm promises to assist in tackling various challenges, including understanding the formation and regulation of mesoscale assemblies inside cells.

Addresses
[1] Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 653, Santiago 7800003, Chile

[2] Department of Physics and Astronomy, Clemson University, Clemson 29634, United States

Corresponding author: Sanabria, Hugo (hsanabr@clemson.edu)

Current Opinion in Structural Biology 2021, **66**:129–138

This review comes from a themed issue on **Folding and binding**

Edited by **Margaret S Cheung** and **Vic Arcus**

For a complete overview see the Issue and the Editorial

Available online 24th November 2020

https://doi.org/10.1016/j.sbi.2020.10.016

Introduction

From sequence to ensemble(s) and function(s)

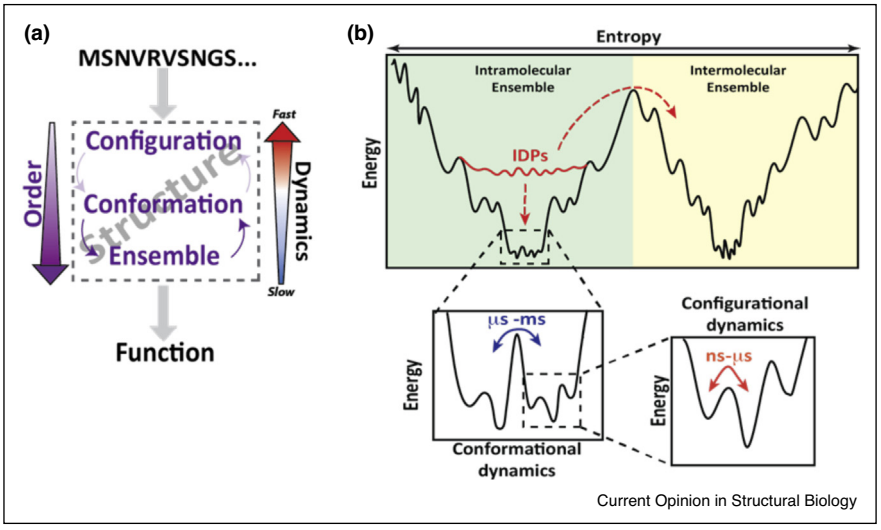
Proteins have a widespread and crucial role in maintaining the cell's metabolism, impacting almost every metabolic process to ensure survival and evolution. Since the first three-dimensional description of a protein at atomic resolution obtained by Perutz in 1960 [1], the unanimated vision of a protein has led to the well-known sequence-to-structure-to-function paradigm. However, the classical textbook paradigm largely disregards the relevance of dynamical processes between these conformations; due to their inherent thermal fluctuations and chain flexibility, proteins continuously explore different configurations, reaching the accumulation of structural conformations that define the whole structural ensemble (Figure 1a). Then, dynamics not only allow reproducing what we observe as the native state, but also characterizes the proteins' function(s), properties, and regulation.

Using his seminal experimental findings, Anfinsen indirectly described the first evidence of inherent dynamics in the folding of a protein [2]. Building on that view, Levinthal [3], and later Wolynes [4], suggested that specific topological constraints from the amino acid chain must guide folding to satisfy the timescales typically observed *in vitro* and *in vivo*. The minimally-frustrated nature of proteins allows them to rapidly explore several short-lived configurations with high structural entropy and low transition energy barriers. Increasing the energetic barriers between configurations causes the adoption of a native/functional ensemble, highlighting the delicate balance between dynamics, structure, and function.

However, the discovery of proteins that show complex folding pathways leading to intricate functions has suggested a revision of this sequence-to-structure-to-function paradigm. Such is the case, for example, of proteins that dimerize via three-dimensional domain swapping (3D-DS) [5••]. These proteins contain local intrinsically disordered regions (IDRs), causing them to lack a well-defined, stable, and minimally-frustrated native ensemble. Moreover, several others are entirely disordered (IDPs) [6–10], showing highly-dynamical competing configurations (Figure 1b). While well-folded proteins show slower transitions as they jump over high energy barriers between distinct states, IDPs must be analyzed at shorter timescales to sample their different configurations due to their faster configurational dynamics.

Interestingly, for most locally or completely disordered proteins [11,12,13•], binding offers a mechanism for folding [14,15], adding a regulatory layer. For binding reactions in proteins and other macromolecules (i.e. nucleic acids) [16,17•,18,19], dynamics can exhibit dominant effects on association and/or dissociation rates by

Figure 1



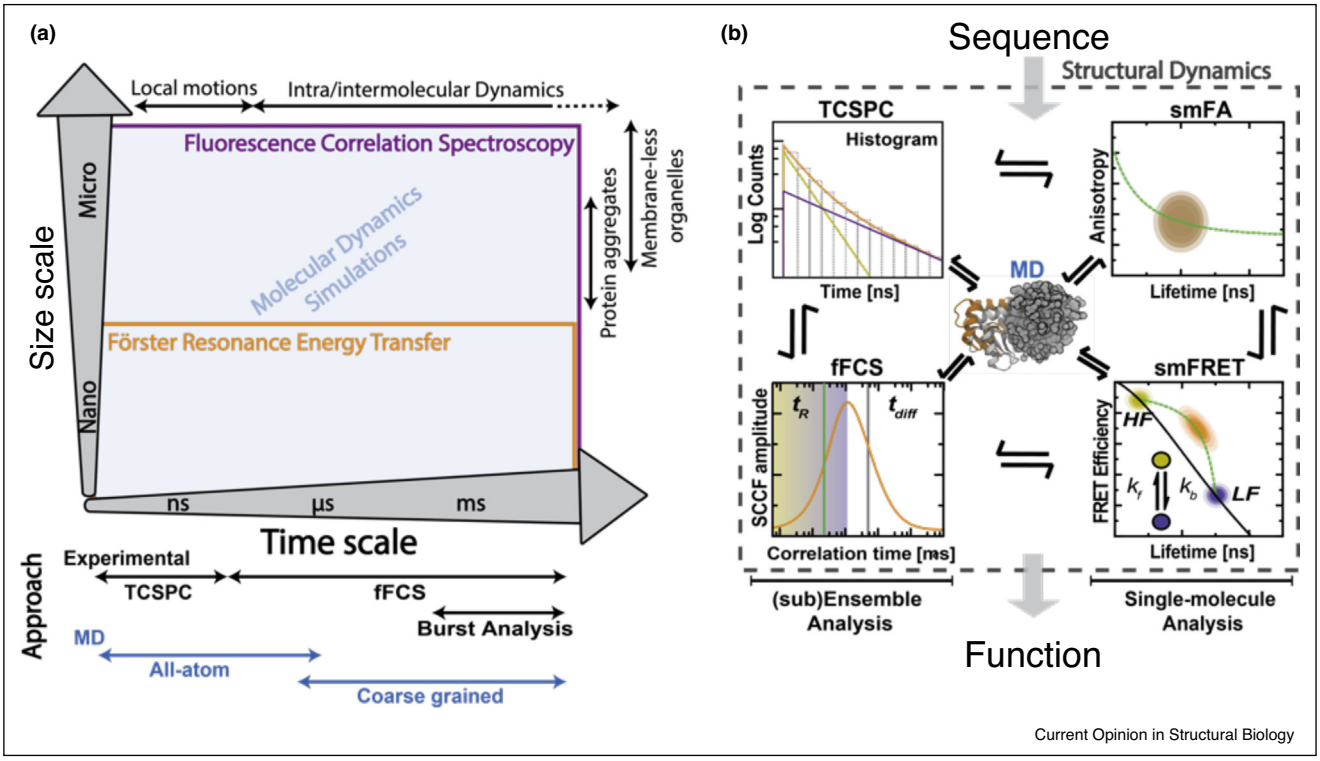
Relationship between sequence, structure, dynamics, and function. **(a)** As the original paradigm stated, the linear sequence-structure-function relationship that explained the proteins' properties and functions obscure the relevance of the structural dynamics. Several meta-stable configurations are grouped into more stable conformations, defining the structural ensemble and what we observe as the native state. **(b)** Depending on the energy barriers between different structural transitions, different dynamic processes can occur between nanoseconds to milliseconds timescales depending on the energy barrier to pass. These processes conform to both intra or intermolecular ensembles. For intrinsically-disordered proteins (IDPs) (red line), the high structural heterogeneity leads to energy frustration. This heterogeneity can be decreased by employing contacts with small ligands or even other macromolecules, adopting intra- or inter-molecular structural ensembles (dashed red lines).

performing a pivotal role in specificity/promiscuity [20–22], thus affecting the lifetime of those complexes. In these cases, binding reactions allow switching between ensembles (Figure 1b). Additionally, binding, (un)folding, and dynamics can modulate micro and mesoscale molecular assemblies, such as membrane-less organelles [23,24•] and liquid phase condensates [25–27,28•], critical components in compartmentalization and other intricate functions within cells (Figure 1b). This new understanding of proteins fills a clear gap in the sequence-to-structure-to-function paradigm to explain numerous biological phenomena where the structure itself is insufficient.

Because solution nuclear magnetic resonance (NMR) can study molecules at the atomic level with a high temporal and spatial resolution [29], it is currently the gold standard ensemble approach to describe local and global structural changes of proteins in folding, binding, and function [30•,31]. As such, NMR gives experimental descriptions of the intra- and intermolecular changes between pico- to milliseconds regimes [32] and dynamic behavior between micro- to milliseconds (and beyond), allowing extensive studies into their involvement in folding and binding [11,33,34] (Figure 2a). However, NMR and other classical ensemble methodologies, although possessing high temporal resolution, struggle to characterize the short-lived configurations of highly-flexible proteins due to the need for high data throughput and ensemble averaging. For IDPs in particular, defined ensembles link to specific functions by integrating and processing signals when folded into stable structures upon binding to cellular regulatory partners, emphasizing the complexity of the (un)folding and function relationship.

The unique advantage of single-molecule methodologies is in their ability to unravel structural heterogeneity, in most cases, without ensemble averaging. Experimental results based on fluorescence are widely exploited due to their excellent structural and temporal resolution [35–37]. Taking into advantage the different approaches and experimental corrections derived from fluorescence, single-molecule multiparameter fluorescence spectroscopy (smMFS) is a robust methodology to accurately monitor and quantify local and global dynamic changes [38–40,41•,42•]. When combined with (sub)ensemble approximations, such as Fluorescence Correlation Spectroscopy (FCS) and Time-Correlated Single-Photon Counting (TCSPC), smMFS allow the monitoring of structural changes in a broad time scale from nano-to-milliseconds [38–40,41•,42•]. For slower processes, approaches focused on fixed molecules are ideal, monitoring real-time structural changes [41•]. Specifically, single-molecule fluorescence anisotropy (smFA) allows the monitoring of local changes that reflect side-chain dynamics. Also, single-

Figure 2

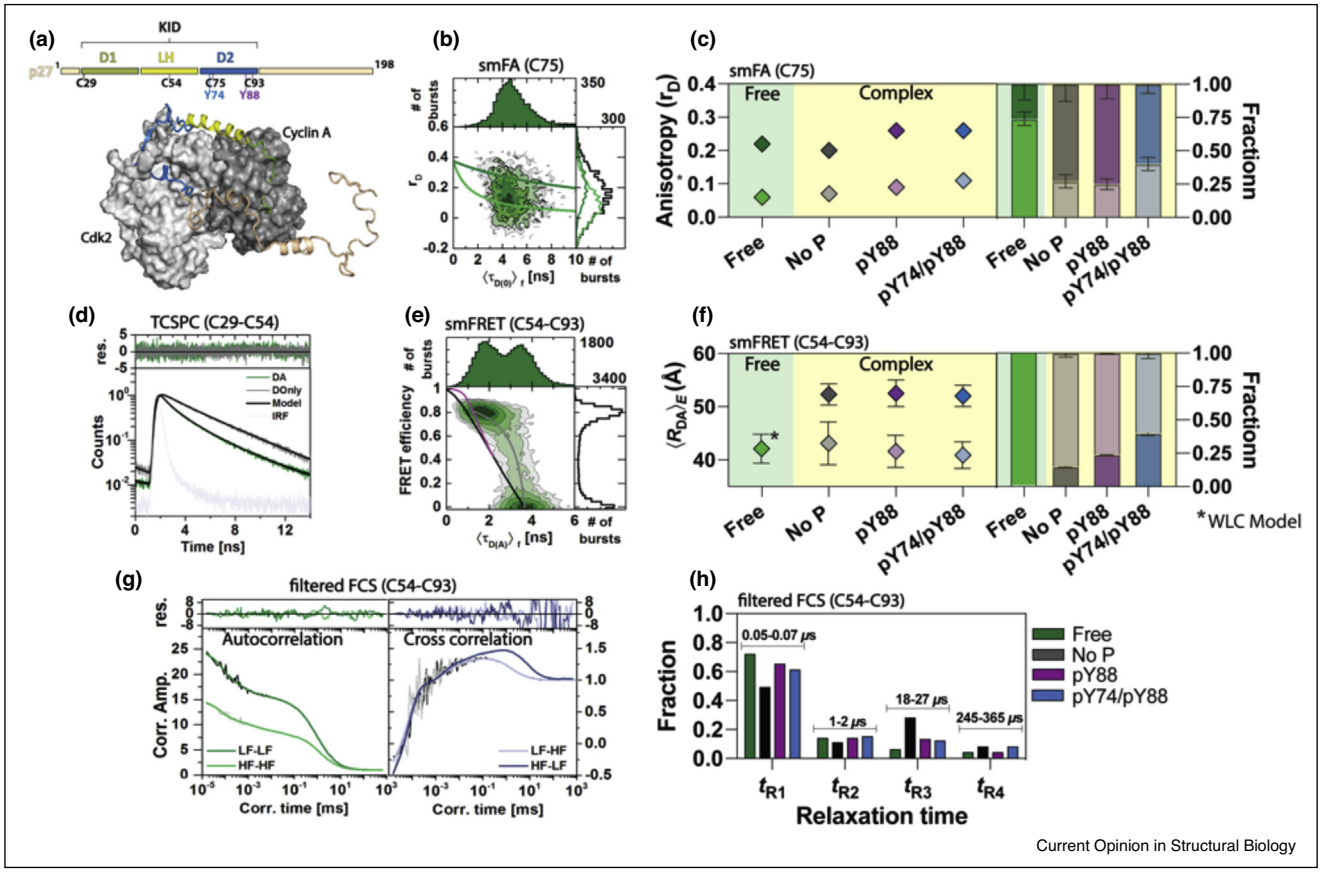


Conformational dynamics and experimental approaches to study. **(a)** Temporal and size scales covered by the combination of experimental and bioinformatic approaches. Main fluorescent methods in freely diffusing conditions (Fluorescence Correlation Spectroscopy -FCS- and Förster Resonance Energy Transfer -FRET-) are used by employing Time-Correlated Single Photon Counting (TCSPC), filtered FCS (fFCS) and burst analysis. These approaches are combined with Molecular Dynamics simulations (MD) (all-atom and coarse-grained) to cover from nano- to milliseconds in temporal resolution and from nano- to micrometers in size scale. For slower temporal scales, microscopy approaches focused on analyzing fixed molecules are ideal. **(b)** Flow chart to study structural dynamics using single-molecule multiparameter fluorescence spectroscopy (smMFS) toolbox. (Sub)ensemble (TCSPC and fFCS) and single-molecule (FRET and anisotropy) approaches, combined with MD can describe local structural changes at high temporal resolution. Each technique provides complementary information to each other, painting a complete picture across the accessible timescales: **TCSPC:** distribution of conformations present in a specified condition (monitored by FRET) that are stable on the nanoseconds timescale (> fluorophore lifetime); **fFCS:** solving of different relaxation times accounting for structural changes across time; **smFA:** high sensitivity to local flexibility changes; **smFRET:** quantification of different distance changes spanning a protein or protein complex via High FRET (HF) or Low FRET (LF) states, distributions of these distances, and the kinetic forward (k_f) and backward (k_f) rates of exchange; **MD:** refinement of structural models generated by the experimental considerations.

molecule Förster Resonance Energy Transfer (smFRET), when used with time-resolved fluorescence spectroscopy (TCSPC), probes distance changes and population heterogeneity with nanoseconds resolution. When coupled to burst analysis, smFRET is sensitive to dynamics over broad time scales, from milliseconds to seconds depending on instrumentation [38–40,41•,42•]. Finally, filtered FCS (fFCS) becomes ideal for following exchange processes between FRET states to quantify the structural dynamics between (sub)micro- to milliseconds (Figure 2b).

Moreover, due to the comparable timescales covered by smMFS and molecular dynamics (MD) simulations (Figure 2a), the combination of experimental results with MD lead to more accurate structural dynamics models to fully understand protein dynamics at the atomic scale [43••]. In particular, coarse-grain models, by requiring less computational resources than all-atom models to manage intra- and intermolecular interactions [44–46], is preferred when modeling larger, complex, multi-protein structures, such as quinary protein structures [47,48]. As such, coarse-grain models have become instrumental in recent modeling [49–53]. By coupling the smMFS with computational approaches, the smMFS toolbox is built (Figure 2b). This toolbox allows us to monitor several aspects of protein function, including folding [36,54], super tertiary [55•,56–59] and quaternary communications [5••,13•], and enzyme catalysis [60••,61], emphasizing how those processes create more extensive, dynamic, three-dimensional systems responsible for life.

Figure 3



smMFS toolbox to study intra- and intermolecular dynamics of p27. **(a)** the topology of human p27 showing the relevant regions (D1, LH, and D2) of its Kinase Inhibitory Domain (KID). C29, C54, C75, and C93 are cysteine residues used to attach the different fluorophores, whereas Y74 and Y88 are tyrosine residues that can be phosphorylated. The ternary complex p27/Cdk2/Cyclin 2 is in cartoon. **(b)** smFA plot for free p27 monitoring C75 attached with Bodipy Fl, showing the two anisotropy population of the D2 region. **(c)** Quantitative analysis of free p27 and the ternary complex with its different phosphorylation modifications (No P, pY88, and pY74/p88), showing anisotropy values and their fraction in all conditions. **(d)** TCSPC plot showing fluorescence decay of free p27 monitoring C29-C54 attached with donor and acceptor of FRET (DA), donor only labeled p27 (DOnly), and the instrument response function (IRF). Differences between DA and DOnly serves as a baseline for comparison and FRET efficiency determination. **(e)** smFRET distribution monitoring distance changes in free p27 labeled, as mentioned in (D). The black line corresponds to the static FRET line, gray line corresponds to the dynamic FRET between DOnly and high FRET, and in pink line, a worm-like chain (WLC) model considering an equilibrium between a disordered and folded protein. **(f)** Quantitative analysis of (E) shows distances and fractions in free p27 in the same conditions as mentioned in (C). For free p27, FRET distance was determined using a WLC model. **(g)** fFCS plots show both auto- and cross-correlation between low and high FRET. **(h)** Quantitative analysis from data obtained in (F) for p27 in all before mentioned conditions. Data fitting found four different exchanging times (t_R) for all conditions, showing the specific fraction for each one.

Uncovering the role of dynamics in the sequence-to-structure-to-function paradigm: conformational heterogeneity as pivotal for proteins' functions

One hallmark model highlighting the relevance of dynamics in protein function at high resolution is the human p27. This disordered protein causes cell cycle arrest when binding in a ternary complex with cyclin-dependent kinase (Cdk2) and cyclins (e.g. Cdk2/Cyclin A) [62,63] (Figure 3a). A recent integrative and collaborative work between multiple laboratories revealed how p27 morphs lead to the formation of the p27/ Cdk2/Cyclin A complex. Different constructs of p27 were studied using stopped-flow kinetics and the smMFS toolbox (Figure 2b) to identify the critical events that led to the initiation complex. An intricate combination of intra- and intermolecular dynamics seems to modulate this protein's biological function (Figure 3).

Intramolecular dynamics: structural heterogeneity of proteins as functional limiting events

Unbound or free p27 is mostly disordered while maintaining some residual alpha-helical structure in the LH subdomain consistent with prior studies [64–67]. As shown in Tsytlonok *et al.* [68••], free p27 adopts a compact conformation, impeding the acquisition of the ternary complex with Cdk2/Cyclin A. Hence, p27 must expand

to expose its 12 residues recognition site in the D1 subdomain [69], being crucial in the association kinetics to Cyclin A by undergoing conformational rearrangement before initial binding [68••]. Similarly, the D2 region must exchange conformations for the recognition of the Cdk2 binding site. Local dynamics monitored by smFA of free p27 (Figure 3b and c) showed that free p27 shows two anisotropy values (Figure 3b and c) that reflects the flexible (low rD) and rigid (high rD) conformations. However, in the absence of target complexes, this protein is preferentially compacted (Figure 3c).

Additionally, by analyzing time-resolved fluorescence (Figure 3d) and single-molecule FRET (Figure 3e) probing various regions of p27, it was found that p27 must expand to create the ternary complex. For example, the FRET variant monitoring dynamics of regions LH and D2 (cysteines for labeling at locations C54 and C93, Figure 3a) shows a dynamic system by which p27 behaves as an intrinsically disordered protein (Figure 3e) following a worm-like-chain (WLC) model with an averaged donor-acceptor distance of 41.8 ± 2.3 Å (magenta dynamic line in Figure 3f). This result is consistent with NMR measurements, MD of the full-length p27, analytical ultracentrifugation, and small-angle- X-ray scattering of the p27/Cdk2/Cyclin A complex [70,71••]. The disordered nature of p27 permits jumps over low energy barriers and rapidly sample multiple configurations that can, over longer timescales, transition between distinct conformations or eventually accessing different structural ensembles, referred as ensemble switching [72,73]. fFCS can efficiently identify all these structural changes over a broad temporal domain (nano-to-milliseconds), corroborating that most of dynamical exchange occurs in the nanoseconds regime (Figure 3g).

Finally, the information derived from discrete MD (DMD) simulations, which samples the heterogeneous landscape, was used as an integrative element in the smMFS toolbox [74–80]. By using radius of gyration (R_g) and α-helical content on the same regions monitored by experimental observations, authors could compare interdye distances, local flexibility and polymeric behavior (like the persistence length). Thus, DMD and smMFS help each other as independent and complementary approaches without imposing physical constrains that biased either simulations of experimental observables into the attained results.

Although very useful for IDP models, this smMFS toolbox is not restricted to highly flexible proteins, but has identified transient conformations even in well-folded and minimally-frustrated models. Using the smMFS toolbox, Sanabria *et al.* [60••] determined the conformations of the lysozyme of bacteriophage T4 (T4L) in the catalytic cycle progression. Three major conformations that are present in the free (E), enzyme-substrate complex (ES), and enzyme-product (EP) bound states. These conformations exchange at few microseconds and hundreds of microseconds, extending the Michalis-Menten mechanism and highlighting that specific conformations favor the progression of the enzymatic reaction. In contrast, for free p27, the transitions observed imply high conformational heterogeneity and flexibility according to its disordered nature (Figure 3h), which suggests that, although disordered, p27 must overcome an expansion to bind with Cdk2/Cyclin A. These examples highlight the relevance of using smMFS toolbox to temporally characterize the structural dynamics of diverse proteins.

Intermolecular dynamics: structural dynamics in multi-step binding and partial dissociation as function modulator

Once defined that p27 must extend to bind the Cdk2/Cyclin A complex, authors studied the main changes involved in forming the ternary complex. Using smFA (Figure 3c), NMR, and X-ray crystallography (X-ray), Tsytlonok *et al.* [71••] discovered that p27 mostly adopts the extended conformation when it is bound to Cdk2/Cyclin A complex. Additionally, by analyzing different donor-acceptor combinations, two limiting states were obtained for p27 in complex. For example, using the FRET variant C54-C93 (covering the LH-D2 regions), authors found two distances with ÅR_{DA} $e_{E,exp}$ that go from 43.1 ± 0.1 to 52.3 ± 0.1 Å via smFRET (Figure 3f), showing a good agreement with the crystallographic structure (PDBID 1JSU). When modeling the accessible volume (AV) of the dyes in such configuration using coarse-grained simulations, results showed experimental-simulations differences within ∼3 Å. The anterior indicates expansion from a more compact conformation to a conformation that exposes the D2 region and adds robust stabilization in the structural dynamics, as observed in fFCS (Figure 3h) by the accumulation of transitions fraction in the mid-microseconds regime. In summary, a fully formed, fuzzy ternary complex built with a simultaneous extension of p27 was identified [68••,71••].

Furthermore, once p27 is bound to Cdk2/Cyclin A and causes cell cycle arrest, this ternary complex is finely regulated via phosphorylation of two occluded tyrosine residues by tyrosine kinases Bcr-Abl and Src for Y88 and Y74 (Figure 3a), respectively [81,82]. For these residues to be phosphorylated through dynamic anticipation, p27 exchanges between different conformations in the bound complex allow the sequential exposure of Y88, followed by Y74 anticipating phosphorylation [71••]. Each of these phosphorylation conditions allow the accessibility of different conformational ensembles. The process was observed by using the smMFS toolbox (Figure 2b) and integrating other biochemical and biophysical methods, including NMR, isothermal titration calorimetry (ITC), and X-ray crystallography.

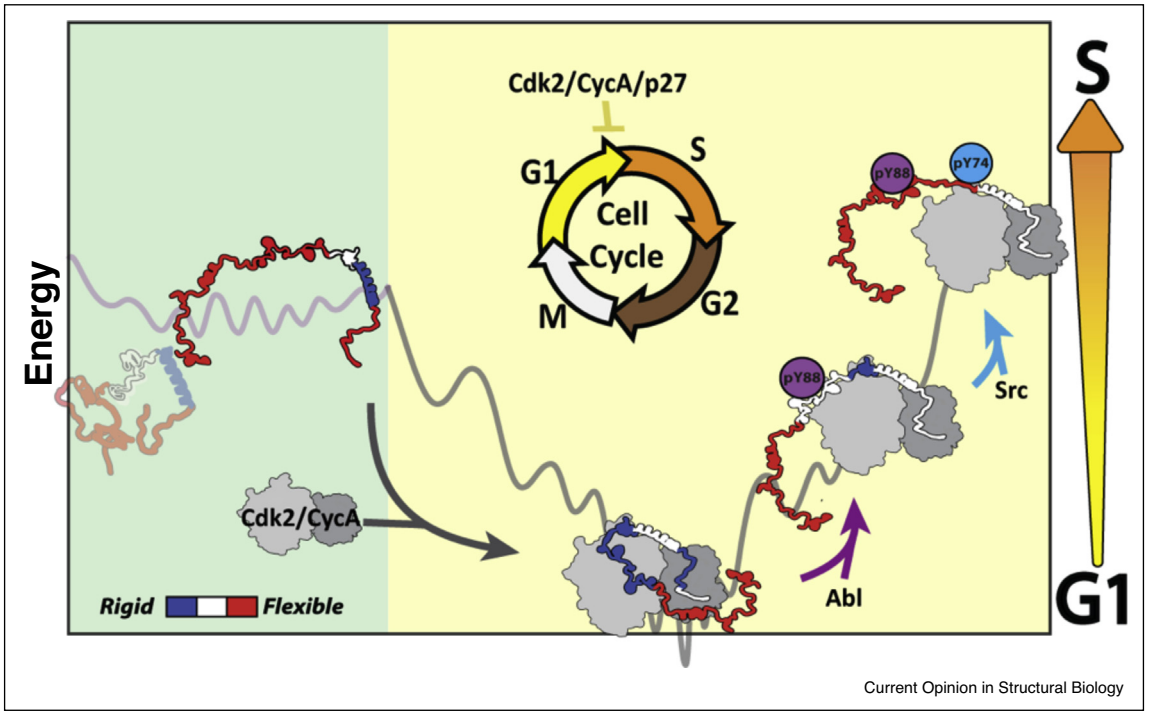
To start, smFA (Figure 3c) showed the release of Y88 followed by Y74, supported by the increase in a population with low anisotropy values, which indicates a more freely rotating fluorophore and in agreement with chemical shift assignments of the D2 domain (e.g., C75 and C93). Next, using the same FRET variant, C54-93, smFRET showed a redistribution of states occurs, shifting the population to a more extended partially released state, thus exposing the phosphorylated Y88 (Figure 3f). In this new state, Y74 is anticipated to be sequentially phosphorylated, evidenced by the release of C75 in smFA after Y88 and Y74 are phosphorylated (Figure 3c). With fFCS, a redistribution towards the accumulation of nanoseconds fraction exchange is described, suggesting that phosphorylation allows the adoption of a highly dynamic p27 is formed [71••].

To showcase the role of partial dissociation and disorder in the structural dynamics, Medina et al. [5••] studied the domain-swapped dimer of the DNA-binding domain of human FoxP1. The compact and folded dimer adopted via 3D-DS exchanges with an extended dimeric, mostly disordered, intermediate ensemble adopting heterogeneous structural changes occurring between 20 μs to 5 ms. The extended intermediate is kinetically allowed due to a low average energetic barrier of ∼1 kcal mol^{-1}, resulting in the intermediate to become highly accumulated as the unfolding of the protein is promoted. This result indicates that the monomer-dimer transition overcomes the characteristic high energy barrier of three-dimensional domain swapping by containing IDRs [83,84]. Overall, the smMFS toolbox is powerful in capturing complex regulatory mechanisms from multi-step binding processes and complex folding pathways, supporting the need for updates of the current unanimated sequence-to-structure-to-function paradigm to a sequence-to-dynamics-to-function.

Perspective: from structural dynamics to function and protein assemblies

For cells to function correctly, proteins must work synergistically. Only by understanding how structural dynamics guide ensemble switching, we can understand how proteins self-assemble into multi-functional three-dimensional mesoscale architectures. Therefore, by following the relationship between dynamics and function, insights can be gained for various genetic diseases such as cancer [85–88], Huntington's [89], autism [90], spinal muscular dystrophy [91].

Figure 4



Functional connection between structural dynamics of p27 and cell cycle. The adoption of the ternary complex between p27/Cdk2/Cyclin A depends on the conformational exchange of free p27 and different conformational ensembles. The expansion of this protein enables the binding to Cdk2/Cyclin A, leading the cell cycle to arrest in phase G, impeding the development of phase S, and therefore DNA replication. Intrinsic flexibility of p27 allows phosphorylation modifications in its tyrosine residues 74 and 88, increasing the expansion and the release from the complex, and the consequent recruitment of ubiquitination proteins that finally leads to the degradation. These events allow Cdk2/Cyclin to continue their functional role in ensuring the cell cycle progression, therefore cell division.

The previous example of p27 binding with Cdk2/Cyclin A to form a ternary complex shows how intra- and intermolecular and intermolecular interactions must work together to regulate the cell cycle [62,63]. Binding only occurs due to the intramolecular behavior of p27, which allows a rapid sampling of multiple configurations to access the extended conformation (Figure 4). Intermolecular interactions with Cdk2/ Cyclin A impedes the cell division by arresting the cycle in phase G. This p27/Cdk2/ Cyclin A association has enormous metabolic significance. Further, it is tightly regulated by specific phosphorylation modifications that trigger p27 ubiquitination followed by degradation. Degradation of p27 enables cell cycle progression [70,81]. However, as discovered, all these events inherently depend on the structural dynamics that characterize p27. The p27/Cdk2/Cyclin A complex is a clear example of where dynamics lead to a change in conformations and defined ensembles that allows the complex to adapt specific functionality. This model and others [56,92,93] have recently revealed the extreme relevance of conformational dynamics as a key functional modulator.

The next logical step is understanding high-order assemblies and their role in modulating the function of the cells. So far, there are few characterized examples by which high-order complexes communicate in relevant processes [58,94]. Such is the case of the dynamics of chromatin, where nucleosome opening/closing transitions stability can severely influence the gene expression activity inside the nucleus. A combination of single-molecule approaches with molecular dynamic simulations found that binding with external proteins severely influences nucleosome dynamics [40,41•,57,95•,96], pivotal to decipher how gene expression occurs. Dynamics are also an essential part of polyfunctional molecules, where molecular adaptors must be coordinated to ensure the appropriate function depending upon the situation [97].

Future studies are required for highly dynamic and less ordered complex systems, such as biomolecular condensates, mitotic spindles, and focal adhesions [98•]. All characterized examples focus the essential role of heterogeneity in dynamics, by which molecules may explore various conformational ensembles, each with crucial consequences in those complexes and their stability. However, although much is still left to understand micro- and mesoscale assemblies within cells, current studies are focused on applying all these high-resolution approaches inside cells to increase the understanding of structural dynamics and assemblies in a real biological context [99,100]. In the near future, we anticipate that this holistic toolbox presented will continue to unravel the sequence-to-function relationship of many mesoscale assemblies in live cells.

Conflict of interest statement

Nothing declared.

Author contributions

EM and DL wrote the outline with guidance from HS. All authors contributed to writing the manuscript.

Acknowledgements

This work acknowledges support from Clemson University, NSF (CAREER MCB- 1749778) and US National Institutes of Health (NIH) P20GM121342. EM was partially supported by a pilot program for global engagement of Clemson University, and by Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt grant 1170701).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest

1. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC: **Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. Resolution, obtained by X-ray analysis**. *Nature* 1960, **185**:416-422.

2. Anfinsen C: **Principles that govern the folding of protein chains**. *Science* 1973, **181**:223-230.

3. Levinthal C: **Are there pathways for protein folding?** *J Chim Phys* 2017, **65**:44-45.

4. Wolynes PG, Onuchic JN, Thirumalai D: **Navigating the folding routes**. *Science* 1995, **267**:1619-1620.

5. •• Medina E, Villalobos P, Hamilton GL, Komives EA, Sanabria H, Ramirez-Sarmiento CA, Babul J: **Intrinsically disordered regions of the DNA-binding domain of human FoxP1 facilitate domain swapping**. *J Mol Biol* 2020, **432**:5411-5429.
In this work, authors get insights about the flexibility and dynamics of the domain-swapped dimer of FoxP1. Using a hybrid between experimental approaches such as smFRET, smFRET-FCS and hydrogen-deuterium exchange with mass spectrometry, combined with molecular dynamics simulations, they showed that dimeric FoxP1 is mostly a dynamic system containing at least two locally-extended intermediates, that facilitates the transitions between the monomer and the domain-swapped dimer.

6. Faust O, Grunhaus D, Shimshon O, Yavin E, Friedler A: **Protein regulation by intrinsically disordered regions: a role for subdomains in the IDR of the HIV-1 rev protein**. *Chembiochem* 2018, **19**:1618-1624.

7. Protter DSW, Rao BS, Van Treeck B, Lin Y, Mizoue L, Rosen MK, Parker R: **Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly**. *Cell Rep* 2018, **22**:1401-1412.

8. Wright PE, Dyson HJ: **Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm**. *J Mol Biol* 1999, **293**:321-331.

9. Wright PE, Dyson HJ: **Intrinsically disordered proteins in cellular signalling and regulation**. *Nat Rev Mol Cell Biol* 2015, **16**:18-29.

10. Uversky VN: **What does it mean to be natively unfolded?** *Eur J Biochem* 2002, **269**:2-12.

11. Miskei M, Horvath A, Vendruscolo M, Fuxreiter M: **Sequence-based prediction of fuzzy protein interactions**. *J Mol Biol* 2020, **432**:2289-2303.

12. Tompa P, Fuxreiter M: **Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions**. *Trends Biochem Sci* 2008, **33**:2-8.

13. • Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, Sottini A, Soranno A, Buholzer KJ, Nettels D *et al.*: **Extreme disorder in an ultrahigh-affinity protein complex**. *Nature* 2018, **555**:61-66.
This works beautifully shows how the binding specificity and affinity do not rely exclusively in binding interfaces but also are present in disordered systems. Using a combination between ensemble, single-molecule approaches, and molecular dynamics simulations, authors found that

the extremely high affinity histone H1/prothymyosin-α complex retains the intrinsics disorder that characterizes both proteins.

14. Zosel F, Mercadante D, Nettels D, Schuler B: **A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction**. *Nat Commun* 2018, **9**:3332.

15. Kim JY, Chung HS: **Diverse transition paths of coupled binding and folding of intrinsically disordered protein proved by three-color single-molecule FRET**. *Biophys J* 2020, **118**:491a-492a.

16. Holmstrom ED, Liu Z, Nettels D, Best RB, Schuler B: **Disordered RNA chaperones can enhance nucleic acid folding via local charge screening**. *Nat Commun* 2019, **10**:2453.

17. • Mitra J, Ha T: **Streamlining effects of extra telomeric repeat on telomeric DNA folding revealed by fluorescence-force spectroscopy**. *Nucleic Acids Res* 2019, **47**:11044-11056.

In this work, authors studied the effect of G-quadruplexes (GQs)—structures observed when high-repeats of T_2AG_3 stack together—stability and dynamics, to understand telomeric single-stranded tails properties due their high GQs content. By combining smFRET and optical twezeers, they found that sequences having five or six T_2AG_3 repeats preferentially fold into GQs by the 3' end and show extreme mechanical and structual stability.

18. Lee HT, Sanford S, Paul T, Choe J, Bose A, Opresko PL, Myong S: **Position-dependent effect of guanine base damage and mutations on telomeric G-quadruplex and telomerase extension**. *Biochemistry* 2020, **59**:2627-2639.

19. Lee CY, McNerney C, Ma K, Zhao W, Wang A, Myong S: **R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation**. *Nat Commun* 2020, **11**:3392.

20. Dogan J, Gianni S, Jemth P: **The binding mechanisms of intrinsically disordered proteins**. *Phys Chem Chem Phys* 2014, **16**:6323-6331.

21. Mollica L, Bessa LM, Hanoulle X, Jensen MR, Blackledge M, Schneider R: **Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment**. *Front Mol Biosci* 2016, **3**:52.

22. Zhou HX, Pang X, Lu C: **Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets**. *Phys Chem Chem Phys* 2012, **14**:10466-10476.

23. Banani SF, Lee HO, Hyman AA, Rosen MK: **Biomolecular condensates: organizers of cellular biochemistry**. *Nat Rev Mol Cell Biol* 2017, **18**:285-298.

24. • Dignon GL, Zheng W, Kim YC, Best RB, Mittal J: **Sequence determinants of protein phase behavior from a coarse-grained model**. *PLoS Comput Biol* 2018, **14**:e1005941.

This works introduces a computational framework to study intrinsically disordered proteins and protein assembly in liquid-liquid phase separation using a coarse-grained model, offering the possibility to simulate these systems without the computational resources needed when using all-atom simulations. The authors tested FUS and LAF-1 proteins, observing good correlation between experimental properties and simulations.

25. Brangwynne CP, Tompa P, Pappu RV: **Polymer physics of intracellular phase transitions**. *Nat Phys* 2015, **11**:899-904.

26. Sahli L, Renard D, Sole-Jamault V, Giuliani A, Boire A: **Role of protein conformation and weak interactions on gamma-gliadin liquid-liquid phase separation**. *Sci Rep* 2019, **9**:13391.

27. Lin YH, Song JH, Forman-Kay JD, Chan HS: **Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins**. *J Mol Liquids* 2017, **228**:176-193.

28. • Dignon GL, Zheng W, Best RB, Kim YC, Mittal J: **Relation between single-molecule properties and phase behavior of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2018, **115**:9929-9934.

Here, authors use their recently developed framework (see Ref. 24) to calculate Θ temperature (for coil-to-globule protein transition, T_Θ), the Boyle temperature (T_B) and the critical temperature (for phase separation, T_C) of twenty proteins. They found a linear correlation between these parameters, suggesting that their correlation serve as a prediction of proteins as contributing in liquid-liquid phase separation and membraneless organels.

29. Kleckner IR, Foster MP: **An introduction to NMR-based approaches for measuring protein dynamics**. *Biochim Biophys Acta* 2011, **1814**:942-968.

30. • Adamski W, Salvi N, Maurin D, Magnat J, Milles S, Jensen MR, Abyzov A, Moreau CJ, Blackledge M: **A unified description of intrinsically disordered protein dynamics under physiological conditions using NMR spectroscopy**. *J Am Chem Soc* 2019, **141**:17817-17829.

An excellent and integrated description of intrinsically-disordered proteins study using NMR, highlighting the insights into the conformational changes in crowded solutions as example of complex environmental conditions, describing dynamics up to ten of nanoseconds.

31. Dyson HJ, Wright PE: **Perspective: the essential role of NMR in the discovery and characterization of intrinsically disordered proteins**. *J Biomol NMR* 2019, **73**:651-659.

32. Mayor T, Hacker U, Stierhof YD, Nigg EA: **The mechanism regulating the dissociation of the centrosomal protein C-Nap1 from mitotic spindle poles**. *J Cell Sci* 2002, **115**:3275-3284.

33. Nolting B, Golbik R, Neira JL, Soler-Gonzalez AS, Schreiber G, Fersht AR: **The folding pathway of a protein at high resolution from microseconds to seconds**. *Proc Natl Acad Sci U S A* 1997, **94**:826-830.

34. Chen EH, Lu TT, Hsu JC, Tseng YJ, Lim TS, Chen RP: **Directly monitor protein rearrangement on a nanosecond-to-millisecond time-scale**. *Sci Rep* 2017, **7**:8691.

35. Kim JY, Kim C, Lee NK: **Real-time submillisecond single-molecule FRET dynamics of freely diffusing molecules with liposome tethering**. *Nat Commun* 2015, **6**:6992.

36. Otosu T, Ishii K, Tahara T: **Microsecond protein dynamics observed at the single-molecule level**. *Nat Commun* 2015, **6**:7685.

37. Kondo T, Pinnola A, Chen WJ, Dall'Osto L, Bassi R, Schlau-Cohen GS: **Single-molecule spectroscopy of LHCSR1 protein dynamics identifies two distinct states responsible for multi-timescale photosynthetic photoprotection**. *Nat Chem* 2017, **9**:772-778.

38. Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CAM: **Accurate single-molecule FRET studies using multiparameter fluorescence detection**. *Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*. Elsevier; 2010:455-514 http://dx.doi.org/10.1016/s0076-6879(10)75018-7:. Methods in Enzymology.

39. Peulen TO, Opanasyuk O, Seidel CAM: **Combining graphical and analytical methods with molecular simulations to analyze time-resolved FRET measurements of labeled macromolecules accurately**. *J Phys Chem B* 2017, **121**:8211-8241.

40. Lehmann K, Felekyan S, Kuhnemuth R, Dimura M, Toth K, Seidel CAM, Langowski J: **Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET**. *Nucleic Acids Res* 2020, **48**:1551-1571.

41. • Gansen A, Felekyan S, Kuhnemuth R, Lehmann K, Toth K, Seidel CAM, Langowski J: **High precision FRET studies reveal reversible transitions in nucleosomes between microseconds and minutes**. *Nat Commun* 2018, **9**:4628.

A beautiful description of structural dynamics in a complex system such as nucleosomes at single-molecule level. Using different FRET variants, the authors quantified destabilization dynamics of mononucleosomes by increasing the ionic stength, finding first opening events in the microseconds timescale by which DNA can breath, until the complete nucleosome dissasembly that is slower than 10 000 seconds.

42. • Craggs TD, Sustarsic M, Plochowietz A, Mosayebi M, Kaju H, Cuthbert A, Hohlbein J, Domicevica L, Biggin PC, Doye JPK *et al.*: **Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase**. *Nucleic Acids Res* 2019, **47**:10788-10800.

This work elegantly describes the structural changes of DNA required to be recognized by the DNA polymarase I, combining smFRET data from 73 variants and molecular dynamics simulations. The authors found that single-nucleotide-gapped DNA bending and fraying are necessary for the polymerase recognition and binding.

43. •• Dimura M, Penlen TO, Sanabria H, Rodnin D, Hemmen K, Hanke CA, Seidel CAM, Gohlke H: **Automated and optimally FRET-assisted structural modeling**. *Nat Commun* 2020 http://dx.doi.org/10.1038/s41467-020-19023-1.
A powerful combination of smFRET and molecular dynamics simulations is used to create a computational suite to allow the modelling of complex biomolecular systems. By using smFRET, a set of distances acts as a structural guide to perform specific molecular dynamics simulations and thererfore obtain a refined model that considers the inherent dynamics.

44. Noid WG: **Perspective: coarse-grained models for biomolecular systems**. *J Chem Phys* 2013, **139** 090901.

45. Devane R, Shinoda W, Moore PB, Klein ML: **A transferable coarse grain non-bonded interaction model for amino acids**. *J Chem Theory Comput* 2009, **5**:2115-2124.

46. Tozzini V: **Coarse-grained models for proteins**. *Curr Opin Struct Biol* 2005, **15**:144-150.

47. Cohen RD, Pielak GJ: **Electrostatic contributions to protein Quinary structure**. *J Am Chem Soc* 2016, **138**:13139-13142.

48. Wirth AJ, Gruebele M: **Quinary protein structure and the consequences of crowding in living cells: leaving the test-tube behind**. *Bioessays* 2013, **35**:984-993.

49. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T: **Valence and patterning of aromatic residues determine the phase behavior of prion-like domains**. *Science* 2020, **367**:694-699.

50. Harmon TS, Holehouse AS, Rosen MK, Pappu RV: **Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins**. *eLife* 2017, **6**.

51. Guin D, Gruebele M: **Weak chemical interactions that drive protein evolution: crowding, sticking, and quinary structure in folding and function**. *Chem Rev* 2019, **119**:10691-10717.

52. Chowdhury A, Bollinger JA, Dear BJ, Cheung JK, Johnston KP, Truskett TM: **Coarse-grained molecular dynamics simulations for understanding the impact of short-range anisotropic attractions on structure and viscosity of concentrated monoclonal antibody solutions**. *Mol Pharm* 2020, **17**:1748-1756.

53. Kozyrev SV, Volovich IV: **Quinary lattice model of secondary structures of polymers**. *Phys A-Stat Mech Appl* 2014, **393**:86-95.

54. Yoo J, Louis JM, Gopich IV, Chung HS: **Three-color single-molecule FRET and fluorescence lifetime analysis of fast protein folding**. *J Phys Chem B* 2018, **122**:11702-11720.

55. • Yanez Orozco IS, Mindlin FA, Ma J, Wang B, Levesque B, Spencer M, Rezaei Adariani S, Hamilton G, Ding F, Bowen ME *et al.*: **Identifying weak interdomain interactions that stabilize the supertertiary structure of the N-terminal tandem PDZ domains of PSD-95**. *Nat Commun* 2018, **9**:3724.
This integrative and collaborative work uses smFRET and molecular dynamics simulations to solve structural interdomain dynamics of the PSD-95, finding that specific conformations described as open and closed that match with previous structure determination. These conformations are exchanging with a low energy barrier between them and are stabilized by specific side-chain interactions.

56. Dahiya V, Agam G, Lawatscheck J, Rutz DA, Lamb DC, Buchner J: **Coordinated conformational processing of the tumor suppressor protein p53 by the Hsp70 and Hsp90 chaperone machineries**. *Mol Cell* 2019, **74**:816-830 e817.

57. Kameda T, Awazu A, Togashi Y: **Histone tail dynamics in partially disassembled nucleosomes during chromatin remodeling**. *Front Mol Biosci* 2019, **6**:133.

58. Gudkov AT, Behlke J, Vtiurin NN, Lim VI: **Tertiary and quaternary structure for ribosomal protein L7 in solution**. *FEBS Lett* 1977, **82**:125-129.

59. Peulen T-O, Hengstenberg CS, Biehl R, Dimura M, Lorenz C, Valeri A, Ince S, Vopel T, Farago B, Gohlke H *et al.*: **Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase**. *arXiv: Biol Phys* 2020.

60. •• Sanabria H, Rodnin D, Hemmen K, Peulen TO, Felekyan S, Fleissner MR, Dimura M, Koberling F, Kuhnemuth R, Hubbell W *et al.*: **Resolving dynamics and function of transient states in single enzyme molecules**. *Nat Commun* 2020, **11**:1231.
Using a combination of smFRET, smFA, fFCS and molecular simulations, a short-lived configurational state of lyzome T4L was found to compliment the already known open and close state. With this discovery, an extended Michaelis-Menten mechanism for enzymes is explained and supported.

61. Sielaff H, Singh D, Grüber G, Börsch M: *Analyzing Conformational Changes in Single FRET-labeled A1 Parts of Archaeal A1AO-ATP synthase*. SPIE; 2018.

62. Polyak K, Lee MH, Erdjument-Bromage H, Koff A, Roberts JM, Tempst P, Massague J: **Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals**. *Cell* 1994, **78**:59-66.

63. Toyoshima H, Hunter T: **p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21**. *Cell* 1994, **78**:67-74.

64. Bienkiewicz EA, Adkins JN, Lumb KJ: **Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1)**. *Biochemistry* 2002, **41**:752-759.

65. Otieno S, Grace CR, Kriwacki RW: **The role of the LH subdomain in the function of the Cip/Kip cyclin-dependent kinase regulators**. *Biophys J* 2011, **100**:2486-2494.

66. Otieno S, Kriwacki R: **Probing the role of nascent helicity in p27 function as a cell cycle regulator**. *PLoS One* 2012, **7**:e47177.

67. Bowman P, Galea CA, Lacy E, Kriwacki RW: **Thermodynamic characterization of interactions between p27(Kip1) and activated and non-activated Cdk2: intrinsically unstructured proteins as thermodynamic tethers**. *Biochim Biophys Acta* 2006, **1764**:182-189.

68. •• Tsytlonok M, Hemmen K, Hamilton G, Kolimi N, Felekyan S, Seidel CAM, Tompa P, Sanabria H: **Specific conformational dynamics and expansion underpin a multi-step mechanism for specific binding of p27 with Cdk2/Cyclin A**. *J Mol Biol* 2020, **432**:2998-3017.
This indept study of p27 uses smFRET, smFA, fFCS, stopped-flow kinetics, and replica-exchange discrete molecular dynamics simulations to determine the multi-step mechanism used to bind p27 to the Cdk2/Cyclin A complex. It was found that the binding mechanism depends on local and long-range configurations.

69. Russo AA, Jeffrey PD, Patten AK, Massague J, Pavletich NP: **Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex**. *Nature* 1996, **382**:325-331.

70. Galea CA, Nourse A, Wang Y, Sivakolundu SG, Heller WT, Kriwacki RW: **Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1**. *J Mol Biol* 2008, **376**:827-838.

71. •• Tsytlonok M, Sanabria H, Wang Y, Felekyan S, Hemmen K, Phillips AH, Yun MK, Waddell MB, Park CG, Vaithiyalingam S *et al.*: **Dynamic anticipation by Cdk2/Cyclin A-bound p27 mediates signal integration in cell cycle regulation**. *Nat Commun* 2019, **10**:1676.
This study shows the importance of protein flexiblity for signaling and function by studying the phosphorylation of p27 bound to Cdk2/Clyclin A. Using NMR spectroscopy, smFA, smFRET and fFCS it was determined that Y88 is first phosphorylated which allows access for Y74 to be phosphorylated. Isothermal titration calorimetry were then used to monitor how tryosine phsophorylations allow the dissasembly of the p27/Cdk2/Cyclin A complex to promote cell division.

72. Choi UB, Sanabria H, Smirnova T, Bowen ME, Weninger KR: **Spontaneous switching among conformational ensembles in intrinsically disordered proteins**. *Biomolecules* 2019, **9**:114.

73. Choi UB, McCann JJ, Weninger KR, Bowen ME: **Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins**. *Structure* 2011, **19**:566-576.

74. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM *et al.*: **Outcome of the first wwPDB hybrid/integrative methods task force workshop**. *Structure* 2015, **23**:1156-1167.

75. Matsunaga Y, Sugita Y: **Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning**. *eLife* 2018, **7**.

76. Dimura M, Peulen TO, Hanke CA, Prakash A, Gohlke H, Seidel CA: **Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems**. *Curr Opin Struct Biol* 2016, **40**:163-185.

77. Ding F, Tsao D, Nie H, Dokholyan NV: **Ab initio folding of proteins with all-atom discrete molecular dynamics**. *Structure* 2008, **16**:1010-1018.

78. Shirvanyants D, Ding F, Tsao D, Ramachandran S, Dokholyan NV: **Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization**. *J Phys Chem B* 2012, **116**:8375-8382.

79. Proctor EA, Ding F, Dokholyan NV: **Discrete molecular dynamics**. *Wiley Interdiscip Rev-Comput Mol Sci* 2011, **1**:80-92.

80. Ding F, Dokholyan NV: **Discrete molecular dynamics simulation of biomolecules**. In *Computational Modeling of Biological Systems*. Edited by Dokholyan NV. Springer US;; 2012:55-73 http://dx.doi.org/10.1007/978-1-4614-2146-7_3. Biological and Medical Physics, Biomedical Engineering.

81. Grimmler M, Wang Y, Mund T, Cilensek Z, Keidel EM, Waddell MB, Jakel H, Kullmann M, Kriwacki RW, Hengst L: **Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases**. *Cell* 2007, **128**:269-280.

82. Chu I, Sun J, Arnaout A, Kahn H, Hanna W, Narod S, Sun P, Tan CK, Hengst L, Slingerland J: **p27 phosphorylation by Src regulates inhibition of cyclin E-Cdk2**. *Cell* 2007, **128**:281-294.

83. Rousseau F, Schymkowitz JW, Itzhaki LS: **The unfolding story of three-dimensional domain swapping**. *Structure* 2003, **11**:243-251.

84. Liu L, Byeon IJ, Bahar I, Gronenborn AM: **Domain swapping proceeds via complete unfolding: a 19F- and 1H-NMR study of the Cyanovirin-N protein**. *J Am Chem Soc* 2012, **134**:4229-4235.

85. Tannapfel A, Grund D, Katalinic A, Uhlmann D, Kockerling F, Haugwitz U, Wasner M, Hauss J, Engeland K, Wittekind C: **Decreased expression of p27 protein is associated with advanced tumor stage in hepatocellular carcinoma**. *Int J Cancer* 2000, **89**:350-355.

86. Li Z, Tao Y, Wang X, Jiang P, Li J, Peng M, Zhang X, Chen K, Liu H, Zhen P *et al.*: **Tumor-secreted exosomal miR-222 promotes tumor progression via regulating p27 expression and re-localization in pancreatic cancer**. *Cell Physiol Biochem* 2018, **51**:610-629.

87. Patel P, Tsiperson V, Gottesman SRS, Somma J, Blain SW: **Dual inhibition of CDK4 and CDK2 via targeting p27 tyrosine phosphorylation induces a potent and durable response in breast cancer cells**. *Mol Cancer Res* 2018, **16**:361-377.

88. Kolsi LE, Leal AS, Yli-Kauhaluoma J, Liby KT, Moreira VM: **Dehydroabietic oximes halt pancreatic cancer cell growth in the G1 phase through induction of p27 and downregulation of cyclin D1**. *Sci Rep* 2018, **8**:15923.

89. Savas JN, Makusky A, Ottosen S, Baillat D, Then F, Krainc D, Shiekhattar R, Markey SP, Tanese N: **Huntington's disease protein contributes to RNA-mediated gene silencing through association with Argonaute and P bodies**. *Proc Natl Acad Sci U S A* 2008, **105**:10820-10825.

90. Gall JG: **Cajal bodies: the first 100 years**. *Annu Rev Cell Dev Biol* 2000, **16**:273-300.

91. Antonellis A, Ellsworth RE, Sambuughin N, Puls I, Abel A, Lee-Lin SQ, Jordanova A, Kremensky I, Christodoulou K, Middleton LT *et al.*: **Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V**. *Am J Hum Genet* 2003, **72**:1293-1299.

92. Heiss G, Ploetz E, Voith von Voithenberg L, Viswanathan R, Glaser S, Schluesche P, Madhira S, Meisterernst M, Auble DT, Lamb DC: **Conformational changes and catalytic inefficiency associated with Mot1-mediated TBP-DNA dissociation**. *Nucleic Acids Res* 2019, **47**:2793-2806.

93. Barth A, Hendrix J, Fried D, Barak Y, Bayer EA, Lamb DC: **Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of Clostridium thermocellum**. *Proc Natl Acad Sci U S A* 2018, **115**:E11274-E11283.

94. Gudkov AT: **The L7/L12 ribosomal domain of the ribosome: structural and functional studies**. *FEBS Lett* 1997, **407**:253-256.

95. • Kilic S, Felekyan S, Doroshenko O, Boichenko I, Dimura M, Vardanyan H, Bryan LC, Arya G, Seidel CAM, Fierz B: **Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1alpha**. *Nat Commun* 2018, **9**:235.

This excelent work combines freely-difussing and fixed smFRET approaches to obtain chromatin dynamics at high temporal and structural resolution. Using three FRET variants, the authors could solve conformational changes from microseconds to seconds with sub nanometer precision, finding not only highly-dynamical stacking interactions between nucleosomes but also that, although compacted by HPF1-α, chromatin fibers maitains their intrinsic dynamics.

96. Mivelaz M, Cao AM, Kubik S, Zencir S, Hovius R, Boichenko I, Stachowicz AM, Kurat CF, Shore D, Fierz B: **Chromatin fiber invasion and nucleosome displacement by the Rap1 transcription factor**. *Mol Cell* 2020, **77**:488-500 e489.

97. Fox SW: **Self-sequencing of amino acids and origins of polyfunctional protocells**. *Orig Life* 1984, **14**:485-488.

98. • Riback JA, Zhu L, Ferrolino MC, Tolbert M, Mitrea DM, Sanders DW, Wei MT, Kriwacki RW, Brangwynne CP: **Composition-dependent thermodynamics of intracellular phase separation**. *Nature* 2020, **581**:209-214.

This scientific milestone gives us a deep understanding about multi-component liquid-liquid phase separation by analyzing intracellular condensates, using the partition coefficient concept, different protein fusions and micrsocopy. By integrating those elements, authors found not only that heterotypic interactions are relevant, but also its interplay with composition-dependent thermodynamics of condensates assemblies inside cells.

99. Heckmeier PJ, Agam G, Teese MG, Hoyer M, Stehle R, Lamb DC, Langosch D: **Determining the stoichiometry of small protein oligomers using steady-state fluorescence anisotropy**. *Biophys J* 2020, **119**:99-114.

100. Coullomb A, Bidan CM, Qian C, Wehnekamp F, Oddou C, Albiges-Rizo C, Lamb DC, Dupont A: **QuanTI-FRET: a framework for quantitative FRET measurements in living cells**. *Sci Rep* 2020, **10**:6504.

Appendix D

Structural Dynamics of Glutamate Signaling Systems by smFRET

Biophysical Journal
Biophysical Perspective

Biophysical Society

Structural Dynamics of Glutamate Signaling Systems by smFRET

Ryan J. Durham,[1] Danielle R. Latham,[2] Hugo Sanabria,[2,*] and Vasanthi Jayaraman[1,*]
[1]University of Texas Health Science Center at Houston, Houston, Texas and [2]Clemson University, Clemson, South Carolina

ABSTRACT Single-molecule Förster resonance energy transfer (smFRET) is a powerful technique for investigating the structural dynamics of biological macromolecules. smFRET reveals the conformational landscape and dynamic changes of proteins by building on the static structures found using cryo-electron microscopy, x-ray crystallography, and other methods. Combining smFRET with static structures allows for a direct correlation between dynamic conformation and function. Here, we discuss the different experimental setups, fluorescence detection schemes, and data analysis strategies that enable the study of structural dynamics of glutamate signaling across various timescales. We illustrate the versatility of smFRET by highlighting studies of a wide range of questions, including the mechanism of activation and transport, the role of intrinsically disordered segments, and allostery and cooperativity between subunits in biological systems responsible for glutamate signaling.

INTRODUCTION

Recent developments in structural biology, mainly through cryogenic electron microscopy (cryo-EM) techniques, have enabled the study of membrane proteins and their complexes faster than previously possible through x-ray crystallography (1–3). These "snapshots" of the proteins in multiple conformations provide a rich background for investigations of the conformational dynamics necessary to understand the mechanisms mediated by these proteins through a multitude of biophysical methods. Methodologies such as nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR), which provide insights into the conformational dynamics, are challenged by the requirement of high concentrations of samples and limitations in protein size. Additionally, NMR and EPR provide a weighted average of the heterogeneous population, causing critical information about the individual dynamics and intermediate configurations to be lost. Single-molecule methods offer a unique solution to ensemble conditions by providing simultaneous structural and kinetics information from proteins in motion. The ability to observe protein conformation as a function of time allows for a direct study of the conformational transitions and kinetics between states (4).

Single-molecule Förster resonance energy transfer (smFRET) has become a mainstream technique for probing biomolecular structural dynamics. As the number of laboratories using smFRET increases, it becomes imperative to create a standardized technique to ensure proper usage. Furthermore, it is essential to ensure that the proper smFRET experimental design is used to answer the question of interest. Förster resonance energy transfer (FRET) uses the nonradiative transfer of energy from a donor fluorophore to a nearby acceptor fluorophore to accurately measure the distance between them (4,5) because the efficiency of the energy transfer depends on the inverse distance between fluorophores to the sixth power (6). However, selecting the appropriate fluorophores and their placement into biomolecules is of crucial importance. Various types of fluorophores have been employed for smFRET, including genetically encoded fluorophores like GFP (7,8) and organic fluorophore molecules like the cyanine dyes (9–11). Additionally, significant work has been done to develop new fluorophores for use in smFRET and to improve the performance of existing fluorophores (11–13). The most critical criteria to consider when selecting a fluorophore are the R_0 factor of the fluorophore pair, how the fluorophore will be attached to the site of interest within the protein, and the size of the fluorophore. The R_0 factor represents the distance in Angstroms between two fluorophores, at which those fluorophores experience half maximal energy transfer. Because FRET measurements are most sensitive to changes in distance when they are close to the R_0, the fluorophores being used should have an R_0-value that is close to the distance between the two sites being measured. Another important factor is the method by which

Submitted July 22, 2020, and accepted for publication October 13, 2020.
*Correspondence: hsanabr@clemson.edu or vasanthi.jayaraman@uth.tmc.edu
Editor: Meyer Jackson.
https://doi.org/10.1016/j.bpj.2020.10.009

the fluorophore will be attached to the protein of interest. Strategies for attaching fluorophores in a site-specific manner are discussed below. Additionally, the size of the fluorophore is significant because large fluorophores can experience steric hindrance with the protein of interest. Steric clashes can alter the conformation of the protein being studied and can also affect the relative orientation of the fluorophores, which can cause error in the measurements. Energy transfer depends upon the alignment of the fluorophores' dipoles in relation to one another. This parameter, referred to as the dipolar orientation or κ^2, could be estimated by measuring the fluorescence anisotropy to minimize errors in the determined distances (14). For best practices, determining the κ^2 distribution for each FRET-labeled sample is recommended instead of the assumption of isotropic averaging, leading to the common use of 2/3. Moreover, the use of fluorophores with long linkers attached to proteins favors isotropic averaging compared with short-linker fluorophores, which restrict mobility. Therefore, longer linkers lead to small experimental errors to a maximum of 7% when considering all uncertainties in the measurements (15). Energy transfer is only possible when the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor (5). Once compatible fluorophores are selected, interfluorophore distances from 10 to 136 Å, with up to 3-Å accuracy (15–18), are measurable using mainly two modalities: ratiometric intensity-based methods or time-resolved fluorescence (19). By attaching compatible fluorophores to a protein, the distance between two locations on the protein is measurable at an instant in time. Special care must be taken to ensure that the incorporation of the fluorophores does not alter the protein (20,21). A variety of attachment chemistry exists to link fluorophores to proteins, including click reactions (22), cysteine reactions with electrophiles (23), and biorthogonal azide-alkyne chemistry via introduction of unnatural amino acids (24,25). Most of the time, the selection of the reactive chemistry is imposed by the wild-type sequence of the protein under study. For the purpose of studying membrane proteins, different treatments are compatible with smFRET experiments. Isolated membrane proteins can be solubilized in detergents (26). Alternatively, incorporating proteins into lipid-containing nanodisks (27) or reconstituted vesicles (28,29) can preserve the physiological lipid context of the protein. Live cell measurements are also possible (8,30).

smFRET implementations can be done with freely diffusing molecules or immobilized molecules on surfaces. Furthermore, data can be collected using confocal measurements or widefield imaging (Table 1). Selecting the appropriate combination depends on the question being asked. Typically, with diffusion-based experiments, confocal measurements are taken. In this case, labeled proteins of interest diffuse through the well-defined confocal volume of the observing microscope, and short bursts of photons are observed (31). These types of experiments are ideal when the dynamic behavior of interest is faster than the average traversal time through the confocal volume (Fig. 1). Furthermore, the concentration of fluorescently tagged protein within the sample must be low enough to ensure that only one fluorescent protein is within the confocal volume at a given time. If a higher protein concentration is required, one could fluorescently tag a portion of the proteins. This will help ensure only one FRET pair is present for a given measurement. In the case of immobilized molecules, molecules are linked to a coverslip with the most appropriate chemistries. Readers are encouraged to read (32–34) for more details. When immobilizing molecules, care must be taken to ensure that the fluorophores do not bind nonspecifically to the protein of interest or to the slide surface (10). Then individual molecules are located and measured over time using photon-counting detectors via confocal detection or using cameras (CMOS, EMCCDs) via widefield. These experiments are useful for observing individual molecules over longer temporal timescales. Photobleaching events are usually tracked to assure only a single molecule is observed at any given time.

It is imperative that experiments be designed with the appropriate temporal scale in mind (Fig. 1). There are three significant factors in smFRET experiments that affect the achievable temporal resolution: protein treatment, detection mode, and data analysis (4). It has already been mentioned that proteins can be immobilized or allowed to diffuse freely. After determining which of these methods works best for

TABLE 1 Comparison of Confocal and Widefield for Freely Diffusing Molecules and Immobilized Molecule Measurements

	Confocal		
Molecule Freedom	Diffusing	Immobilized	Widefield
Dynamic ranges	picoseconds to diffusion time	nanoseconds to minutes	submillisecond to minutes
Detectors	confocal, photon-counting modules (i.e., PMTs and APDs with TCSPC)		TIRF (widefield) EMCCDs and CMOSs
Number of molecules	limited diffusion through the sample volume, can be over 100,000	100–500[a]	limited by molecules in widefield, is ~500
Data analysis	trace, burst-wise, BVA, fluctuation spectroscopy, and time resolved		trace, BVA, fluctuation spectroscopy, and camera or image based

APD, avalanche photodiode; PMT, photon-multiplier tube; TIRF, total internal reflection fluorescence.
[a]TCSPC is less practical in immobilized molecules because of the lower number of sample molecules, which leads to lower photon-counting statistics sample size and limits the minimal dynamic range.

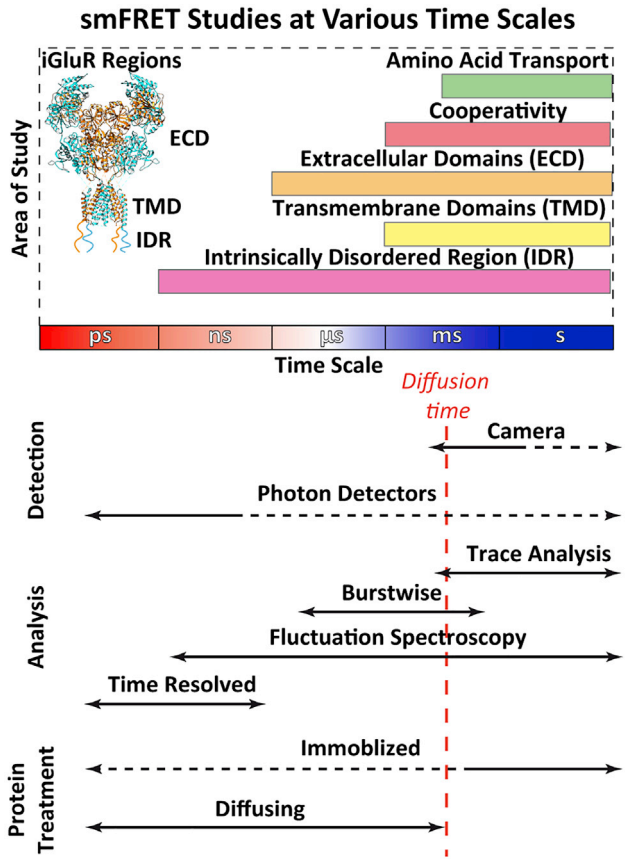


FIGURE 1 smFRET investigations across different timescales. Top: the typical timescales for studies of various areas of glutamate signaling are shown as horizontal bars. The areas of study include both functional areas (e.g., amino acid transport) and structural areas (e.g., the extracellular domain (ECD)). ECD, TMD, and IDR refer to the various domains of the iGluRs as shown in the inset. smFRET investigations of glutamate signaling have focused primarily on the second-to-millisecond timescale, with sub-millisecond studies lagging behind. Bottom: various detection and analysis schemes exist for performing smFRET experiments at a variety of time resolutions. The ranges of timescales, across which the various methods are useful, are shown as horizontal arrows. Solid arrows indicate optimal timescales while dashed arrows indicate sub-optimal time scales. The timescale of diffusion of free-moving molecules is also shown as the timescale of the dynamics of interest can determine whether a diffusion-based or immobilized scheme is more useful for a planned investigation of single-molecule dynamics. To see this figure in color, go online.

making observations at the right timescale for the dynamics of interest, the proper detector must be selected. It is typical to immobilize proteins and use a widefield camera when measuring conformational dynamics that are slower than milliseconds. The photon-to-electron conversion of most widefield camera detectors, such as common charge-coupled devices (CCDs), electron-multiplying CCDs (EMCCDs), and complementary metal-oxide semiconductors (CMOSs), limits the temporal resolution to be around tens of milliseconds (Fig. 1). This temporal resolution is suitable for measuring larger domain motion and large-scale conformational rearrangement of proteins (32). By binning pixels, CMOS cameras can increase their temporal resolution with a time resolution of up to 250 μs being reported (35), but this lowers the spatial resolution. Photon-counting modules like avalanche photodiodes (APDs) and photon multiplier tubes (PMTs) can be utilized to further increase the temporal resolution. Temporal resolution is going to be based on the data-acquisition system used, i.e., Data Acquisition boards (DAQ) and Field-Programmable Gate Arrays (FPGAs). Because these modules cannot differentiate between emission wavelengths, it is crucial to ensure the correct filters are being used based on the selected fluorophores. Photon counting allows the measurement of faster dynamics such as local flexibility and side-chain rotations with a temporal resolution ranging from 10 μs to 100 ms (36), including several orthogonal analysis methods, that increases the extracted dynamics and the temporal information obtained from smFRET measurements (37–39). Time-correlated, single-photon-counting (TCSPC) electronics will measure fluorescence and luminescence lifetimes ranging from nanoseconds to microseconds, respectively (38). Fluorescence correlation spectroscopy (FCS) and its different variations allow conformational dynamics to be determined over a broader timescale (40–43).

There exist many different methods for smFRET data analysis, depending upon the information gathered. Trace analysis is typically modeled with Hidden Markov modeling (HMM) (44,45) and many variations (46). Burst-wise analysis, which looks at the pattern in photon detection to differentiate donor from acceptor in confocal measurements (47), is often done by probability distribution analysis (PDA) (48) or burst variance analysis (BVA) (49). The choice of data analysis depends upon the data collected and experimental design. If FCS and lifetime information is gathered, these data should be included in data analysis because it will only help enhance the results. Multiparameter fluorescence detection (MFD) (19,50) will ensure the best results are obtained. Many laboratories are actively developing new software and setting up best practices in terms of data analysis and offer these packages online for free. These community-based resources range from online software-sharing hubs such as the FRET community (51) and the kinSoftChallenge (46) to multilaboratory studies of FRET precision (15). Given that the size of the protein of interest and the sample concentration are not limitations for FRET experimental design, smFRET is a versatile methodology to study membrane proteins and has been used in a wide range of systems such as G-protein-coupled receptors (GPCRs), transporters, and ion channels. Here, we choose the three systems related to glutamate signaling as an example system in which smFRET methods have provided invaluable biophysical insights. This system was chosen because it has the common theme of glutamate signaling but with three unique systems that covers GPCRs, transporters, and ion channels.

Glutamate signaling systems studied by smFRET

Glutamatergic signaling is the primary form of excitatory signaling in the mammalian central nervous system. In

this type of signaling, the neurotransmitter glutamate is released into the synapse, where it binds to and activates various receptor molecules, including ion channels and metabotropic receptors. After this glutamate release, the glutamate must undergo reuptake into the presynaptic neuron and be packaged into neurotransmitter vesicles, a process facilitated by amino acid transporter molecules (46). smFRET studies have been carried out on proteins from each of these three families of proteins that are involved in glutamatergic signaling. Although the metabotropic glutamate receptor (mGluR) and ionotropic glutamate receptor (iGluR) share some similarities in the glutamate binding site, the transmembrane segments and signaling processes are very different, one being G-protein coupled and the other an ion channel. The glutamate transport has a completely different architecture both at the ligand binding site and in the transmembrane transporter segments. Thus, although the three systems chosen are linked by the fact that they are involved in glutamate-mediated signaling, the wide range of architectures illustrates the versatility of smFRET as a tool to study the dynamics and conformations of a wide variety of proteins.

One family of proteins that are involved in the glutamatergic signaling process are the mGluRs, which are class C GPCRs with large bi-lobed extracellular domains that bind to the agonist glutamate. Based on end-state structures of agonist- and antagonist-bound forms of the extracellular domain, the mechanism for activation by agonists was suggested to be a closure of the bi-lobed cleft (52–56). smFRET investigations on the soluble extracellular domain that utilized a diffusing experimental setup with confocal detection were able to put dynamics into the context of activation and showed that the domain rapidly fluctuates between the open and closed cleft states and that the extent of stabilization of the protein in the closed cleft state dictated the extent of agonism (Fig. 2 *A*; (57)). Although later smFRET measurements on the full-length receptor that analyzed immobilized molecules using a TIRF-based detection scheme showed three major states when measuring across the dimeric extracellular domain at the same sites as in the soluble domain, agonist efficacy was still shown to be dictated by the occupancy of a closed cleft state that is consistent with the smFRET studies on the isolated extracellular domain (58). However, the dwell times for interconversion between the FRET states were found to be in the range of tens of milliseconds and not as rapid as those observed for the isolated extracellular domains. It is possible that the slower dynamics is due to the stabilization and restricted dynamics of the extracellular domains in the presence of the transmembrane domains, showing the importance of investigating the dynamics in the full-length systems.

smFRET measurements of immobilized molecules were also used to investigate the role that conformational changes play in determining the cooperativity between the agonist binding sites in mGluRs (59). These studies showed that

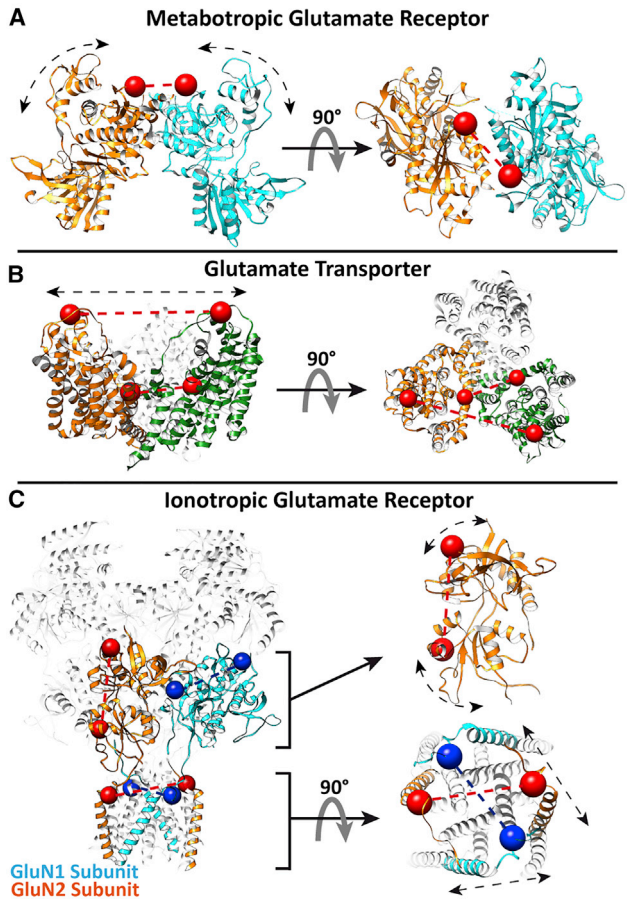


FIGURE 2 smFRET measurements of proteins involved in glutamate signaling. The locations of smFRET fluorophore labeling sites and measurements are shown. These sites are chosen with the goal of studying established (based on previous structures) or hypothesized conformational changes in these proteins. (*A*) Shown are the measurements performed on the agonist-binding dimer of mGluRs (57–60) (Protein Data Bank, PDB: 1EWT (54)). This measurement site reports on the conformational changes at the interface between the mGluR dimer caused by the closure of the bi-lobed cleft upon binding agonists. (*B*) Shown are the measurements performed on the trimeric glutamate transporter (61–63) (PDB: 1XFH (64)). These measurements report on the distance between adjacent monomers of the transporter trimer that are expected to monitor motions associated with transport. (*C*) Shown are the measurements performed on an iGluR (26) (PDB: 6MMK (65)). The sites across the agonist binding domain monitor motions across the bi-lobed cleft caused by agonist or antagonist binding. The sites across the transmembrane segments are expected to monitor motions across the ion pore. The side chains of fluorophore-labeled residues are shown as spheres, whereas distances being investigated are shown as dotted lines. To see this figure in color, go online.

agonist-induced closure of one cleft leads to an allosteric shift in the dynamic equilibrium of the second unliganded subunit in the dimeric mGluRs. Thus, the higher spontaneous basal dynamics and closure of the mGluR3 influenced the liganded mGluR2, leading to higher activation in the heteromer. This mechanism of cooperativity was found to be universal and translated to mGluR2/mGluR7 heteromers, highlighting the importance of such heteromerization in physiology (60). mGluR7 homodimers have a significantly

lower affinity and activation relative to the other mGluRs. However, in the context of mGluR2/mGluR7 heteromers, both efficacy and affinity are high. smFRET studies showed that the heteromeric receptor exhibits faster state-to-state transitions and, therefore, has increased conformational dynamics relative to the homomeric receptor. Additionally, synthetic agonists selective for either mGluR2 or mGluR7 produced a more substantial shift to low FRET (related to higher cleft closure at the agonist binding domain) than observed when both subunits were bound in the corresponding homomeric receptor, correlating well to a stronger activation. These experiments that allow for studying individual heteromeric combinations are possible because specific combinations can be pulled down onto the slides used for smFRET imaging (10,26,66). This selective pulldown of heteromeric receptors is achieved by placing the fluorophore-labeling sites on one subunit and placing the affinity tag for attachment to the slide on the other subunit. Using this strategy, molecules that do not contain a subunit with the affinity tag will not attach to the slide, and molecules that do not contain any fluorophore-labeling sites will not be observed. This ensures that immobilized molecules that exhibit an FRET signal contain at least one of each of these two subunits.

Glutamate transporters catalyze neurotransmitter uptake from the synaptic cleft into the cytoplasm of glial cells and neurons. The bacterial homolog of the glutamate transporter, the sodium-aspartate symporter from *Pyrococcus horikoshii*, Glt_{pH}, has served as a model system for studying structure-function correlations in this family. The initial framework provided by the x-ray structures of the protein in several conformations along with the slow turnover rate of tens of seconds for this protein made it ideal for direct structure-function studies using tethered smFRET (Fig. 2 *B*; (61,62,64,67)). These studies provided direct evidence for "elevator-like" motions for transport and also showed that each subunit could undergo motions independent of each other (28,62,63). This motion also has a burst-like pattern with periods of quiescence and periods of rapid transitions (28). The rapid dynamic mode was hypothesized to be due to the separation of the transport domain from the trimeric scaffold, allowing for the rapid domain movements across the bilayer. Two mutations introduced into Glt_{pH} for imparting characteristics of the human glutamate transporter lead to increased transport domain dynamics. The increased dynamics correlated to the increased rate of substrate transport observed in the human glutamate transporters relative to Glt_{pH}, providing a direct temporal relationship between transport domain motion and substrate uptake (63). Several of these investigations were performed in proteoliposomes, with the transporter tethered on the slide. Tethering allows for a specific outside-out orientation of the transporter, making it possible to tune the concentration of luminal and external ions and substrates, leading to more physiologically relevant studies.

An additional class of membrane proteins that are investigated through smFRET is ion channels. Extensive smFRET investigations have been carried out on members of the iGluRs. Initial studies were performed on the soluble agonist binding domain of the receptor, in particular the AMPA subtype of the iGluRs. X-ray structures of the isolated agonist binding domain showed a correlation between the extent of cleft closure at the bi-lobed agonist binding domain and the extent of activation in several cases (68–73). However, this relationship broke down in mutants, such as the T686S, in which the cleft showed full closure even when the extent of activation was only partial (74). smFRET measurements that utilized immobilized protein molecules on a surface probed all the states that the protein occupies and reconciled this issue by showing that the mutant protein occupied a wide range of conformations and that the activation is dictated by the fractional occupancy of the high FRET closed cleft agonist binding state (75,76). More importantly, the smFRET studies also showed that the distance changes between the most probable states for the different liganded conditions and the mutant correlated with the x-ray structures, showing that distance changes can be accurately measured by smFRET.

Similarly, cryo-EM structures of full-length AMPA receptors showed varying degrees of decoupling across the dimers within the tetrameric amino-terminal domain of the receptor that is associated with receptor desensitization. Hence, the role of this decoupling in desensitization was debated (77–79). The complete conformational landscape that could be probed with smFRET at this site using the immobilized full-length receptor showed that largely decoupled states did exist under desensitizing conditions. However, a larger fraction of the receptor showed smaller decoupling, this large decoupling is not required for desensitization (80). Additionally, based on the distances, we could directly relate the conformations observed in the smFRET data to specific cryo-EM structures (80). Recently, smFRET investigations were used to study the mechanism of cooperativity between the agonists glutamate and glycine in the NMDA receptor subtype of the iGluRs (Fig. 2 *C*; (26)). The smFRET investigations of immobilized receptors showed that the binding of one agonist causes a stabilization of the closed cleft bi-lobed agonist binding domain and lower conformational fluctuations at the site where the agonist binds but an increase in conformational flexibility and dynamics at the second agonist site. The loss of such an effect in a mutant receptor where such negative cooperativity was not observed (81) confirmed that certain conformational states observed via smFRET at the second agonist binding site contribute to the lower affinity of the second agonist when the first agonist is bound to the receptor. These studies again highlight the importance of conformational dynamics in function along with the need to understand the complete conformational landscape of the receptor that is possible through single-molecule methods.

smFRET also stands to contribute significantly to our understanding of macromolecular dynamics for intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs). The inherently dynamic nature of IDPs is incompatible with analysis for static structure through cryo-EM or x-ray crystallography, either because the study is unsuited for analyzing a molecule with little structure or because the results fail to capture a complete understanding of the dynamics of the protein. However, smFRET can report on the conformation of a protein over time without averaging over an ensemble. Combining static measurements and smFRET allows the gathering of valuable information concerning IDPs or IDRs (82,83). For example, from studies of both diffusing and immobilized proteins, it was determined that the IDR C-terminal domain (CTD) of the GluN2B subunit is required for GluN2B to regulate NMDARs (57,60,84–86). The use of smFRET has helped to understand the role of posttranslational modifications on the iGluRs' disordered intracellular segments (87–89). Bowen and Choi used smFRET measurements of immobilized molecules to show differences in conformational dynamics as well as changes in the extent of disorder due to phosphorylation (87,90). Given that there is little structural insight for this segment of the protein, these early studies pave the way for future investigations into how disorder and dynamics in these segments modulate receptor function.

Perspective

To observe several essential conformations for mediating functions of glutamate singling systems, smFRET must be used. This allows for determining the complete conformational landscape and dynamics that are not observed using only structural methods, such as cryo-EM and x-ray crystallography. smFRET also highlights the dynamic nature of these systems and the role of such dynamics in function. The smFRET technique's versatility in terms of the ability to probe protein complexes of all sizes with minimal requirements in terms of concentration and in a near-native state allows for similar investigations on the large number of membrane proteins for which end-state structures are available through cryo-EM. Moreover, with the ability to measure FRET with high precision, it is now possible to use FRET-derived distances in combination with other structural biology tools or molecular dynamics (MD) simulations to model protein structures (91,92) and deposit them in pdb-dev (93). Static structures enable the rational design of starting points for MD simulations and the selection of fluorophore attachment points for smFRET measurements. MD simulations show dynamic motions not revealed with static structures alone and reveal the regions of a molecule that will show biologically relevant motions that can be measured with smFRET. smFRET measurements can provide constraints for MD simulations that enhance the physiological relevance of simulation results and provide insight into highly dynamic regions of proteins that cannot be well resolved through static structural methods alone. However, it is important to understand all the physical aspects and parameters of the dye to obtain accurate simulations. As machine learning continues to advance, it will become a great tool for smFRET data analysis (94,95).

Furthermore, the field of correlative microscopy is yet to expand on the possibility of combining FRET and cryoEM. In this way, structural biology techniques, MD simulations, and smFRET measurements can be combined to provide a greater understanding of various biological systems on the single-molecule level. The development of detectors and data analysis methods that function at submillisecond time resolutions has enabled the investigation of new research questions relating to the structure and dynamics of individual molecules. However, the studies that utilize smFRET have focused primarily on the timescale of seconds to milliseconds, with submillisecond studies lagging (Fig. 1). The future of the smFRET field will involve expanding investigations into the submillisecond dynamics of individual molecules. As the tools and methods needed to investigate submillisecond dynamics become more widely available, with data analysis methods becoming increasingly standardized, the area of submillisecond dynamics will continue to expand and answer as-yet-unresolved research questions.

AUTHOR CONTRIBUTIONS

R.J.D. wrote the outline with guidance from H.S. and V.J. All authors contributed to writing the manuscript.

ACKNOWLEDGMENTS

Funding was provided by National Institutes of Health grants R35 GM122528 to V.J. and F31GM130035 to R.J.D. and National Science Foundation CAREER MCB1749778 and National Institutes of Health 1P20GM121342 to H.S.

REFERENCES

1. Nwanochie, E., and V. N. Uversky. 2019. Structure determination by single-particle cryo-electron microscopy: only the sky (and intrinsic disorder) is the limit. *Int. J. Mol. Sci.* 20:4186.
2. Callaway, E. 2020. Revolutionary cryo-EM is taking over structural biology. *Nature.* 578:201.
3. Rawson, S., S. Davies, …, S. P. Muench. 2016. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Mol. Membr. Biol.* 33:12–22.
4. Lerner, E., T. Cordes, …, S. Weiss. 2018. Toward dynamic structural biology: two decades of single-molecule Förster resonance energy transfer. *Science.* 359:eaan1133.
5. Forster, T. 1946. Energiewanderung und Fluoreszenz. *Naturwissenschaften.* 33:166–175.
6. Clegg, R. M. 2006. The history of FRET. *In* Reviews in Fluorescence 2006. C. D. Geddes and J. R. Lakowicz, eds. Springer, pp. 1–45.
7. Shin, Y., J. H. Davis, …, M. J. Lang. 2009. Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc. Natl. Acad. Sci. USA.* 106:19340–19345.

8. Murakoshi, H., R. Iino, …, A. Kusumi. 2004. Single-molecule imaging analysis of Ras activation in living cells. *Proc. Natl. Acad. Sci. USA.* 101:7317–7322.
9. Cannon, B., J. Kuhnlein, …, T. T. Paull. 2013. Visualization of local DNA unwinding by Mre11/Rad50/Nbs1 using single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 110:18868–18873.
10. Dolino, D. M., S. Chatterjee, …, V. Jayaraman. 2017. The structure-energy landscape of NMDA receptor gating. *Nat. Chem. Biol.* 13:1232–1238.
11. Zheng, Q., M. F. Juette, …, S. C. Blanchard. 2014. Ultra-stable organic fluorophores for single-molecule research. *Chem. Soc. Rev.* 43:1044–1056.
12. Dave, R., D. S. Terry, …, S. C. Blanchard. 2009. Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophys. J.* 96:2371–2381.
13. Altman, R. B., D. S. Terry, …, S. C. Blanchard. 2011. Cyanine fluorophore derivatives with enhanced photostability. *Nat. Methods.* 9:68–71.
14. VanDerMeer, B. W. 2020. Kappaphobia is the elephant in the fret room. *Methods Appl. Fluoresc.* 8:030401.
15. Hellenkamp, B., S. Schmid, …, T. Hugel. 2018. Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nat. Methods.* 15:669–676.
16. Stryer, L. 1978. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* 47:819–846.
17. Stryer, L., and R. P. Haugland. 1967. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA.* 58:719–726.
18. Baibakov, M., S. Patra, …, J. Wenger. 2019. Extending single-molecule förster resonance energy transfer (FRET) range beyond 10 nanometers in zero-mode waveguides. *ACS Nano.* 13:8469–8480.
19. Sisamakis, E., A. Valeri, …, C. A. M. Seidel. 2010. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* 475:455–514.
20. Tyagi, S., and E. A. Lemke. 2015. Single-molecule FRET and crosslinking studies in structural biology enabled by noncanonical amino acids. *Curr. Opin. Struct. Biol.* 32:66–73.
21. Voith von Voithenberg, L., and D. C. Lamb. 2018. Single pair Förster resonance energy transfer: a versatile tool to investigate protein conformational dynamics. *BioEssays.* 40:1700078.
22. Nikić, I., J. H. Kang, …, E. A. Lemke. 2015. Labeling proteins on live mammalian cells using click chemistry. *Nat. Protoc.* 10:780–791.
23. Brustad, E. M., E. A. Lemke, …, A. A. Deniz. 2008. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* 130:17664–17665.
24. Lee, T. C., M. Kang, …, A. A. Deniz. 2016. Dual unnatural amino acid incorporation and click-chemistry labeling to enable single-molecule FRET studies of p97 folding. *ChemBioChem.* 17:981–984.
25. Noren, C. J., S. J. Anthony-Cahill, …, P. G. Schultz. 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science.* 244:182–188.
26. Durham, R. J., N. Paudyal, …, V. Jayaraman. 2020. Conformational spread and dynamics in allostery of NMDA receptors. *Proc. Natl. Acad. Sci. USA.* 117:3839–3847.
27. Nath, A., A. J. Trexler, …, E. Rhoades. 2010. Single-molecule fluorescence spectroscopy using phospholipid bilayer nanodiscs. *Methods Enzymol.* 472:89–117.
28. Akyuz, N., R. B. Altman, …, O. Boudker. 2013. Transport dynamics in a glutamate transporter homologue. *Nature.* 502:114–118.
29. Fitzgerald, G. A., D. S. Terry, …, S. C. Blanchard. 2019. Quantifying secondary transport at single-molecule resolution. *Nature.* 575:528–534.
30. Sako, Y., S. Minoghchi, and T. Yanagida. 2000. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* 2:168–172.
31. Gopich, I. V., and A. Szabo. 2007. Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B.* 111:12925–12932.
32. Jain, A., R. Liu, …, T. Ha. 2011. Probing cellular protein complexes using single-molecule pull-down. *Nature.* 473:484–488.
33. Roy, R., S. Hohng, and T. Ha. 2008. A practical guide to single-molecule FRET. *Nat. Methods.* 5:507–516.
34. Choi, U. B., K. R. Weninger, and M. E. Bowen. 2012. Immobilization of proteins for single-molecule fluorescence resonance energy transfer measurements of conformation and dynamics. *Methods Mol. Biol.* 896:3–20.
35. Juette, M. F., D. S. Terry, …, S. C. Blanchard. 2016. Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nat. Methods.* 13:341–344.
36. Chung, H. S., and I. V. Gopich. 2014. Fast single-molecule FRET spectroscopy: theory and experiment. *Phys. Chem. Chem. Phys.* 16:18644–18657.
37. Wozniak, A. K., G. F. Schröder, …, F. Oesterhelt. 2008. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA.* 105:18337–18342.
38. Dolino, D. M., S. Rezaei Adariani, …, H. Sanabria. 2016. Conformational selection and submillisecond dynamics of the ligand-binding domain of the N-methyl-d-aspartate receptor. *J. Biol. Chem.* 291:16175–16185.
39. Laurence, T. A., X. Kong, …, S. Weiss. 2005. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. USA.* 102:17348–17353.
40. Nettels, D., A. Hoffmann, and B. Schuler. 2008. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J. Phys. Chem. B.* 112:6137–6146.
41. Magde, D., E. Elson, and W. W. Webb. 1972. Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* 29:705–708.
42. Elson, E. L., and D. Magde. 1974. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers.* 13:1–27.
43. Magde, D., W. W. Webb, and E. L. Elson. 1978. Fluorescence correlation spectroscopy. III. Uniform translation and laminar flow. *Biopolymers.* 17:361–376.
44. Okamoto, K. 2017. Analyzing single molecule FRET trajectories using HMM. *In* Hidden Markov Models, *Volume 1552*. D. Westhead and M. Vijayabaskar, eds.. Humana Press, pp. 103–113.
45. McKinney, S. A., C. Joo, and T. Ha. 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* 91:1941–1951.
46. Schmid, S., and M. Götz. 2019. kinSoftChallenge 2019. https://sites.google.com/view/kinsoftchallenge/home.
47. Hagai, D., and E. Lerner. 2019. Systematic assessment of burst impurity in confocal-based single-molecule fluorescence detection using brownian motion simulations. *Molecules.* 24:2557.
48. Antonik, M., S. Felekyan, …, C. A. M. Seidel. 2006. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B.* 110:6970–6978.
49. Torella, J. P., S. J. Holden, …, A. N. Kapanidis. 2011. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* 100:1568–1577.
50. Kudryavtsev, V., M. Sikor, …, D. C. Lamb. 2012. Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements. *Chemphyschem.* 13:1060–1078.
51. Barth, A., T. Peulen, …, C. Seidel. 2020. FRET Community. https://www.fret.community.
52. Bessis, A.-S., P. Rondard, …, J. P. Pin. 2002. Closure of the Venus flytrap module of mGlu8 receptor and the activation process: insights from mutations converting antagonists into agonists. *Proc. Natl. Acad. Sci. USA.* 99:11097–11102.

53. Kniazeff, J., A. S. Bessis, …, J. P. Pin. 2004. Closed state of both binding domains of homodimeric mGlu receptors is required for full activity. *Nat. Struct. Mol. Biol.* 11:706–713.

54. Kunishima, N., Y. Shimada, …, K. Morikawa. 2000. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature.* 407:971–977.

55. Tsuchiya, D., N. Kunishima, …, K. Morikawa. 2002. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+. *Proc. Natl. Acad. Sci. USA.* 99:2660–2665.

56. Koehl, A., H. Hu, …, B. K. Kobilka. 2019. Structural insights into the activation of metabotropic glutamate receptors. *Nature.* 566:79–84.

57. Olofsson, L., S. Felekyan, …, E. Margeat. 2014. Fine tuning of sub-millisecond conformational dynamics controls metabotropic glutamate receptors agonist efficacy. *Nat. Commun.* 5:5206.

58. Vafabakhsh, R., J. Levitz, and E. Y. Isacoff. 2015. Conformational dynamics of a class C G-protein-coupled receptor. *Nature.* 524:497–501.

59. Levitz, J., C. Habrian, …, E. Y. Isacoff. 2016. Mechanism of assembly and cooperativity of homomeric and heteromeric metabotropic glutamate receptors. *Neuron.* 92:143–159.

60. Habrian, C. H., J. Levitz, …, E. Y. Isacoff. 2019. Conformational pathway provides unique sensitivity to a synaptic mGluR. *Nat. Commun.* 10:5572.

61. Reyes, N., C. Ginter, and O. Boudker. 2009. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature.* 462:880–885.

62. Erkens, G. B., I. Hänelt, …, A. M. van Oijen. 2013. Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. *Nature.* 502:119–123.

63. Akyuz, N., E. R. Georgieva, …, S. C. Blanchard. 2015. Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature.* 518:68–73.

64. Yernool, D., O. Boudker, …, E. Gouaux. 2004. Structure of a glutamate transporter homologue from Pyrococcus horikoshii. *Nature.* 431:811–818.

65. Jalali-Yazdi, F., S. Chowdhury, …, E. Gouaux. 2018. Mechanisms for zinc and proton inhibition of the GluN1/GluN2A NMDA receptor. *Cell.* 175:1520–1532.e15.

66. Jain, A., R. Liu, …, T. Ha. 2012. Single-molecule pull-down for studying protein interactions. *Nat. Protoc.* 7:445–452.

67. Ciftci, D., G. H. M. Huysmans, …, O. Boudker. 2020. Single-molecule transport kinetics of a glutamate transporter homolog shows static disorder. *Sci. Adv.* 6:eaaz1949.

68. Armstrong, N., and E. Gouaux. 2000. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron.* 28:165–181.

69. Armstrong, N., M. Mayer, and E. Gouaux. 2003. Tuning activation of the AMPA-sensitive GluR2 ion channel by genetic adjustment of agonist-induced conformational changes. *Proc. Natl. Acad. Sci. USA.* 100:5736–5741.

70. Gouaux, E. 2004. Structure and function of AMPA receptors. *J. Physiol.* 554:249–253.

71. Armstrong, N., Y. Sun, …, E. Gouaux. 1998. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature.* 395:913–917.

72. Birdsey-Benson, A., A. Gill, …, D. R. Madden. 2010. Enhanced efficacy without further cleft closure: reevaluating twist as a source of agonist efficacy in AMPA receptors. *J. Neurosci.* 30:1463–1470.

73. Ahmed, A. H., Q. Wang, …, R. E. Oswald. 2009. Structure of the S1S2 glutamate binding domain of GLuR3. *Proteins.* 75:628–637.

74. Robert, A., N. Armstrong, …, J. R. Howe. 2005. AMPA receptor binding cleft mutations that alter affinity, efficacy, and recovery from desensitization. *J. Neurosci.* 25:3752–3762.

75. Landes, C. F., A. Rambhadran, …, V. Jayaraman. 2011. Structural landscape of isolated agonist-binding domains from single AMPA receptors. *Nat. Chem. Biol.* 7:168–173.

76. Ramaswamy, S., D. Cooper, …, V. Jayaraman. 2012. Role of conformational dynamics in α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptor partial agonism. *J. Biol. Chem.* 287:43557–43564.

77. Dürr, K. L., L. Chen, …, E. Gouaux. 2014. Structure and dynamics of AMPA receptor GluA2 in resting, pre-open, and desensitized states. *Cell.* 158:778–792.

78. Meyerson, J. R., J. Kumar, …, S. Subramaniam. 2014. Structural mechanism of glutamate receptor activation and desensitization. *Nature.* 514:328–334.

79. Yelshanskaya, M. V., M. Li, and A. I. Sobolevsky. 2014. Structure of an agonist-bound ionotropic glutamate receptor. *Science.* 345:1070–1074.

80. Carrillo, E., S. A. Shaikh, …, V. Jayaraman. 2020. Mechanism of modulation of AMPA receptors by TARP-γ8. *J. Gen. Physiol.* 152:e201912451.

81. Regalado, M. P., A. Villarroel, and J. Lerma. 2001. Intersubunit cooperativity in the NMDA receptor. *Neuron.* 32:1085–1096.

82. Ferreon, A. C. M., Y. Gambin, …, A. A. Deniz. 2009. Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 106:5645–5650.

83. Gomes, G.-N. W., M. Krzeminski, …, C. C. Gradinaru. 2020. Integrating multiple experimental data to determine conformational ensembles of an intrinsically disordered protein. *bioRxi* https://doi.org/10.1101/2020.02.05.935890.

84. Punnakkal, P., P. Jendritza, and G. Köhr. 2012. Influence of the intracellular GluN2 C-terminal domain on NMDA receptor function. *Neuropharmacology.* 62:1985–1992.

85. Choi, U. B., R. Kazi, …, M. E. Bowen. 2013. Modulating the intrinsic disorder in the cytoplasmic domain alters the biological activity of the N-methyl-D-aspartate-sensitive glutamate receptor. *J. Biol. Chem.* 288:22506–22515.

86. Warnet, X. L., H. Bakke Krog, …, M. Kjaergaard. 2020. The C-terminal domains of the NMDA receptor: how intrinsically disordered tails affect signalling, plasticity and disease. *Eur. J. Neurosci* Published online May 28, 2020. https://doi.org/10.1111/ejn.14842.

87. Choi, U. B., S. Xiao, …, M. E. Bowen. 2011. Effect of Src kinase phosphorylation on disordered C-terminal domain of N-methyl-D-aspartic acid (NMDA) receptor subunit GluN2B protein. *J. Biol. Chem.* 286:29904–29912.

88. Chatterjee, S., C. Ade, …, C. F. Landes. 2019. Phosphorylation induces conformational rigidity at the C-terminal domain of AMPA receptors. *J. Phys. Chem. B.* 123:130–137.

89. Gutzeit, V. A., J. Thibado, …, J. Levitz. 2019. Conformational dynamics between transmembrane domains and allosteric modulation of a metabotropic glutamate receptor. *eLife.* 8:e45116.

90. Lau, C. G., and R. S. Zukin. 2007. NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nat. Rev. Neurosci.* 8:413–426.

91. Yanez Orozco, I. S., F. A. Mindlin, …, H. Sanabria. 2018. Identifying weak interdomain interactions that stabilize the supertertiary structure of the N-terminal tandem PDZ domains of PSD-95. *Nat. Commun.* 9:3724.

92. Dimura, M., T. Peulen, …, H. Gohlke. 2020. Automated and optimally FRET-assisted structural modeling. *Nat. Commun.* 11:5394.

93. Sali, A., H. M. Berman, …, J. D. Westbrook. 2015. Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure.* 23:1156–1167.

94. Thomsen, J., M. B. Sletfjerding, …, N. S. Hatzakis. 2020. DeepFRET: rapid and automated single molecule FRET data classification using deep learning. *bioRxiv* https://doi.org/10.1101/2020.06.26.173260.

95. Matsunaga, Y., and Y. Sugita. 2018. Linking *time-series* of single-molecule experiments with molecular dynamics simulations by machine learning. *eLife.* 7:e32668.

Appendix E

Size-Exclusion Chromatography-Based Determination of RH of GFP

Size-Exclusion Chromatography-Based Determination of R_h of GFP

Write Up by George Hamilton; Major edits by Danielle Latham
REU Site: Nature's machinery through the prism of physics, biology, chemistry and engineering
Clemson University, Clemson SC

Introduction:

In the general sense, chromatography refers to any methodology that is used to separate the component parts of a solution by passing the *mobile phase*, or the mixture to be separated, through some *stationary phase*, or a material with properties that allow separation of each component. Some of the most common kinds of chromatography used in biophysics research include ion-exchange chromatography, hydrophobic interaction chromatography, and ***size-exclusion chromatography.*** The key takeaway is that each kind of chromatography is specialized for separation of solution components a particular parameter—in the cases listed above, their net charges, hydrophobicities, and molecular sizes or weights, respectively. Typically, this is accomplished through use of a specialized column through which the solution of interest is flowed. However, the protocols for these different chromatography experiments can be drastically different from one another.

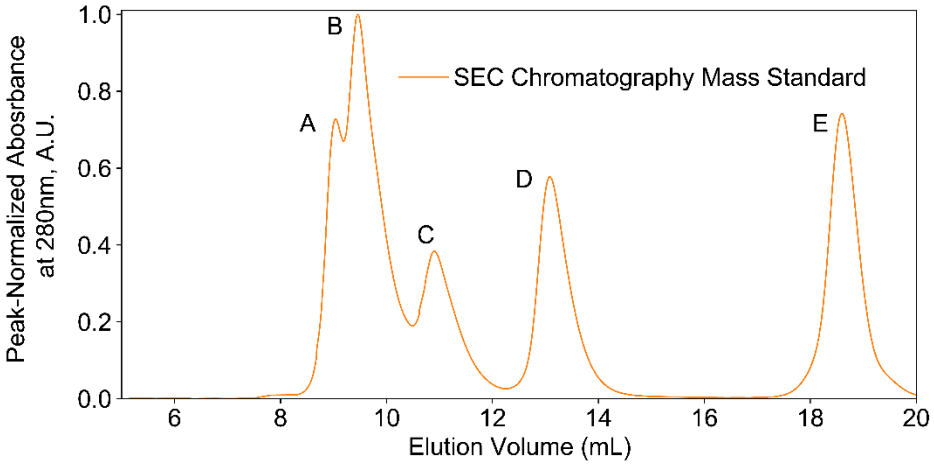
In this exercise, you will perform a complete size-exclusion chromatography experiment to identify green fluorescent protein and calculate its Stokes Radius, or hydrodynamic radius, R_h. We will perform this experiment using a fast protein liquid chromatography (FPLC) setup fitted with an Enrich SEC70 10x300 size-exclusion column. This column has an ideal separation range of 500-70,000 Daltons. Observation of molecules flowing through our column is performed after flowthrough using a UV-vis-spectrometer to observe the degree to which light at certain wavelengths is absorbed by our sample. The column will be equilibrated in the required buffer, then a size-exclusion standard mixture will be ran through the column. The standard mixture contains five molecular species with differing molecular weights and R_h values. These are:

(a) Thyroglobulin (670 kDa, 8.50 nm)
(b) γ-globulin (158 kDa, 5.50 nm)
(c) Ovalbumin (44 kDa, 2.85 nm)
(d) Myoglobin (17 kDa, 1.9 nm)
(e) Vitamin B12 (1.35 kDa, 0.85 nm)

Initial measurement of this standard mixture will provide a context through which we can interpret our data from the unknown mixture containing GFP. An example chromatogram for this standard protein mixture is shown below using observation wavelength of 280 nm. 230 nm and 280 nm absorptions are generally considered standards for identification and

quantification of proteins since 280 nm is the average absorbance maximum for amino acids, and 230 nm is a convenient average for other contaminants.

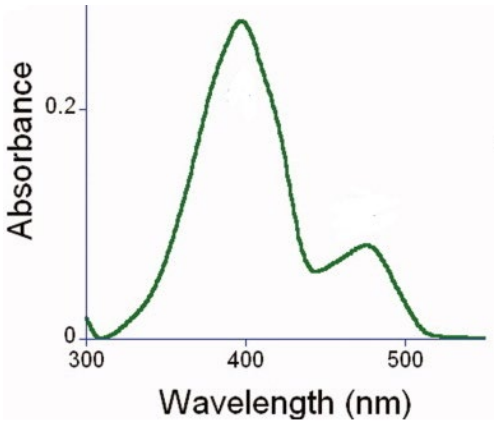
Note that larger, heavier molecules flow through the column first, and lighter molecules flow through last. This is because the size-exclusion column consists of material filled with small pores. Smaller molecules are more likely to be trapped in these pores, whereas larger molecules are less likely. Further, molecules beyond the optimal separation range of 70 kDa (thyroglobulin and γ-globulin) tend to be "squeezed" together. Therefore, our main interest is in the remaining three peaks for standards purposes.



After running our standard, we will run our GFP. It is known that wild-type GFP has a molecular weight of approximately 27 kDa. Where would we expect its peak to be located in the above standards plot? Further, wild-type GFP has an absorption spectrum with a maximum located at 397 nm with a weaker peak near 495 nm, as shown in the figure below.

Absorbance of GFP

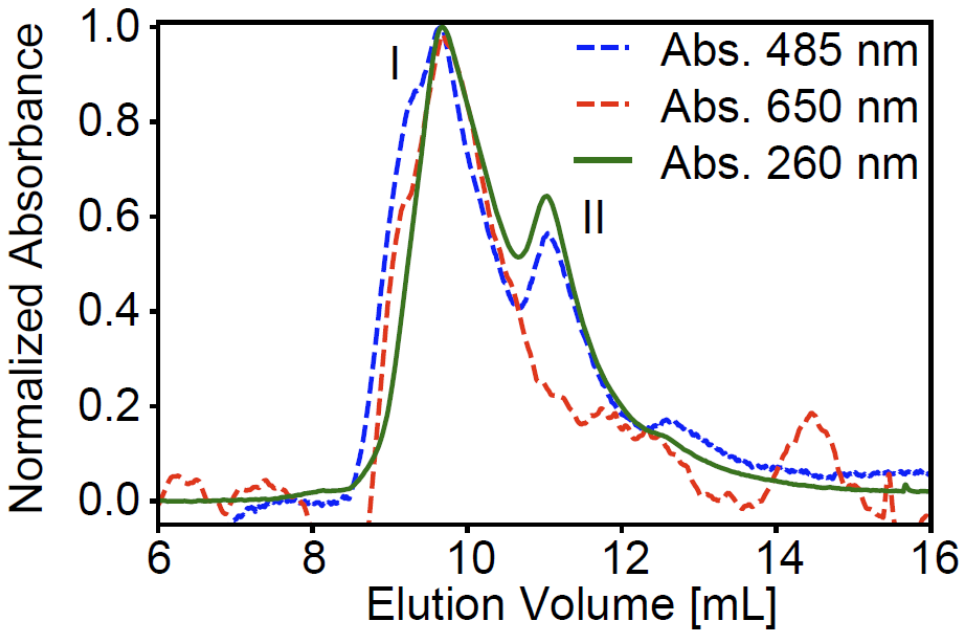
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3190146/



Therefore, if we observe our mixture using our standard lines (280 nm) and GFP's absorption maxima, we should be able to identify GFP's Stoke's Radius.

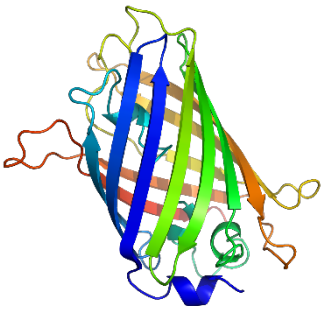
Method Example:

This approach has been previously used to identify fluorescently labeled RNA molecules. The example is shown below. In this case, they were observing at wavelengths corresponding to the RNA maximum at 260 nm as well as absorbance maxima for two fluorescent dyes. Coincidence of all three peaks allowed identification of the labeled molecule. Can you determine the elution volume this occurred at?



So far, we have only seen plots that report elution volume, or the amount of fluid flowed through the column. However, the elution volume is directly related to the R_h of the molecular species corresponding to each peak. This relationship depends on many factors, and so the most direct and accurate way to account for all of them in calculating R_h is to use the standards data to construct a standard line or curve. This curve, in principle, accounts for any small variability in runs through a column that may affect the result. Thus the standard is run in a manner consistent with the solution of interest. Therefore, standards should be run in the same buffer conditions as well as just before experiments with unknown samples. Essentially, we will use the known R_h values and elution volumes for the peaks within the linear separation range from our standard to fit a standard line that directly relates R_h and elution volume. We will then plug in the elution volume for our identified GFP peak into the fitted line and calculate R_h for GFP. To do this, we will plot $\log_{10}(R_h)$ vs. elution volume for the three peaks

corresponding to the SEC70 column's optimal separation range. We will fit this scatter plot with a line of the form y=mx+b. Finally, we will use the fit parameters and plug in the GFP elution volume to get R_h. We will perform identification of peaks during the lab, but calculation of R_h for GFP will be left to each group.



As GFP is heated, it will start to denature and unfold from its typical barrel shape. While denatured, the general structure of GFP will change. We can use the SEC70 column to determine how the change in structure changes the Stokes radius of GFP. When GFP denatures, the structure takes up more space, causing an increase in Stokes radius. How will this change the elution volume?

Further, native GFP will still fluoresce the typical green color while the denatured GFP will not due to the barrel no longer protecting the chromophore within.

Using the SEC70 column and the FPLC method, we will be able to determine the stokes radius of native type and denatured GFP.

Protocol:

- Equilibrate SEC70 column in PBS buffer (This will be done for you)
    - Warm up the absorbance detectors
        - set to wavelengths 230 nm, 260 nm, and 280 nm, corresponding to standard protein absorbance lines.
    - Run H2O at 1mL/min for 2 column volumes (~50 mL)
        - IMPORTANT: H2O must be run to remove storage ethanol and prevent precipitation of salts from buffer
    - Run buffer at 1mL/min for 2 column volumes (~50 mL)
    - Run buffer at .75mL/min until pressure equilibration, at least a few mL
- Run Mass/R_h standard for SEC70 column
    - Standard is pre-diluted into 1mL buffer and stored in 4*C.
    - Loading sample: 50uL
        - Prior to run, load loop will be cleaned/equilibrated. Choose setting for FPLC with sample loop going directly to waste.
        - Unscrew syringe port. Remove 1mL syringe. Load syringe with 1mL H2O using needle (do not load from top by removing plunger). Remove bubble/force to top of syringe by repeated flicking.
        - Insert syringe into syringe port, re-tighten syringe port screw, and gradually inject H2O into sample loop, being sure not to inject air bubbles.

        - Repeat the steps taken for H2O with buffer used for sample and standard dilutions.
        - Finally, remove the syringe once more. Load syringe with 2X sample loop volume (50uL sample loop, so 100uL sample) of standard solution.
        - Remove air as well as possible, re-insert syringe, and load sample into sample loop. USE EXTRA CAUTION to not allow air into sample loop. Sample loop should fill completely, and this should be visible through the loop as standard solution is dark red.
    - Run buffer at 0.75mL/min for 25mL ensuring to include sample loop
    - Save run results for later analysis.
        - We will use the last 3 peaks, corresponding to mass ranges within the optimal linear separation range for our column (<70 kDa), to build a standards curve by which to calibrate our R_h determination of GFP. (I will provide an Excel sheet for identifying peaks in datasets.)
- Run GFP sample
    - GFP sample will be provided in buffer and run as-is without dilution.
    - Zero the absorbance detectors
        - set to wavelengths 280 nm, 397 nm, and 500 nm corresponding to typical protein absorbance, the wild-type Green Fluorescence Protein (GFP) absorbance peak, and a background measure.
    - Clean the sample loop then load 50uL of GFP
    - Run buffer at 0.75mL/min for 25mL ensuring to include sample loop
    - Save data for later analysis.
- Run Heated GFP sample
    - The GFP will need to be heated
        - Take 100uL of GFP in a microcentrifuge tube
        - Using the Isotemp, heat your sample to 70*C for 10 minutes
    - Clean the sample loop then load 50uL of heated GFP
    - Run buffer at 0.75mL/min for 25mL ensuring to include the sample loop
    - Save data for later analysis

Data Analysis:

- Getting the numbers
    - Use Python to:
        - Identify final three peaks from standard chromatogram (the molecules with masses in the linear response range of the column, <70kDA)
        - Plot $\log_{10}(R_h)$ vs elution volume
        - Linear fit. This is the standards calibration curve.

    - Identify GFP peak through combination of chromatogram curves in different wavelengths. The GFP peak should appear in all three observed wavelengths simultaneously.
    - Use fit formula from standard curve to determine R_h for GFP (plug in elution volume, obtain R_h).
    - Check if the obtained value makes sense in terms of the peak location and standard R_h values.
- Write lab report
    - In your own words, describe the methodology used. Describe the setup, the configuration, standard, etc.
    - Consider the advantages and disadvantages of chromatography relative to the other techniques you have used or will use. How much sample is used? How well can you identify GFP? How well does the R_h value obtained agree with other methods and expectations?
    - What kinds of data plots are useful in describing how you arrived at your result?
    - Report your result.

Appendix F

Fluorescence Correlation Spectroscopy (FCS): Clemson's Biophysics REU 2022

Fluorescence Correlation Spectroscopy (FCS)

Clemson's Biophysics REU 2022

Danielle Latham, Rajen Goutam

Background

Introduction

Fluorescence correlation spectroscopy (FCS) is a powerful single-molecule detection technique that measures and correlates fluctuations in fluorescence intensity within a very small detection volume - on the order of femtoliters. The theory of FCS is based on Poisson statistics. This section introduces the physical processes and mathematical theory that govern FCS**.**

FCS is based on analysis of time-dependent intensity fluctuations resulting from dynamic processes, typically translation diffusion into and out of a small volume defined by a focused light beam and a confocal aperture [see figure 1a]. When a fluorescent molecule diffuses into a focused light beam, there is a burst of emitted photons due to multiple excitation-emission cycles from the same molecule. If the molecule diffuses rapidly out of the volume, the photon burst is short-lived. If the molecule diffuses more slowly, the photon burst has a longer duration. By correlation analysis of the time-dependent emission [see figure 1b], one can determine the diffusion coefficient of the molecule. In addition to translation diffusion, intensity fluctuations can occur due to ligand–macromolecule binding, rotational diffusion, internal macromolecule dynamics, intersystem crossing, and excited-state reactions. Correlation functions are used to interpret the collect measurements. Different equations are used to describe each process. Typically, two or more processes occur at the same time, affecting the data.

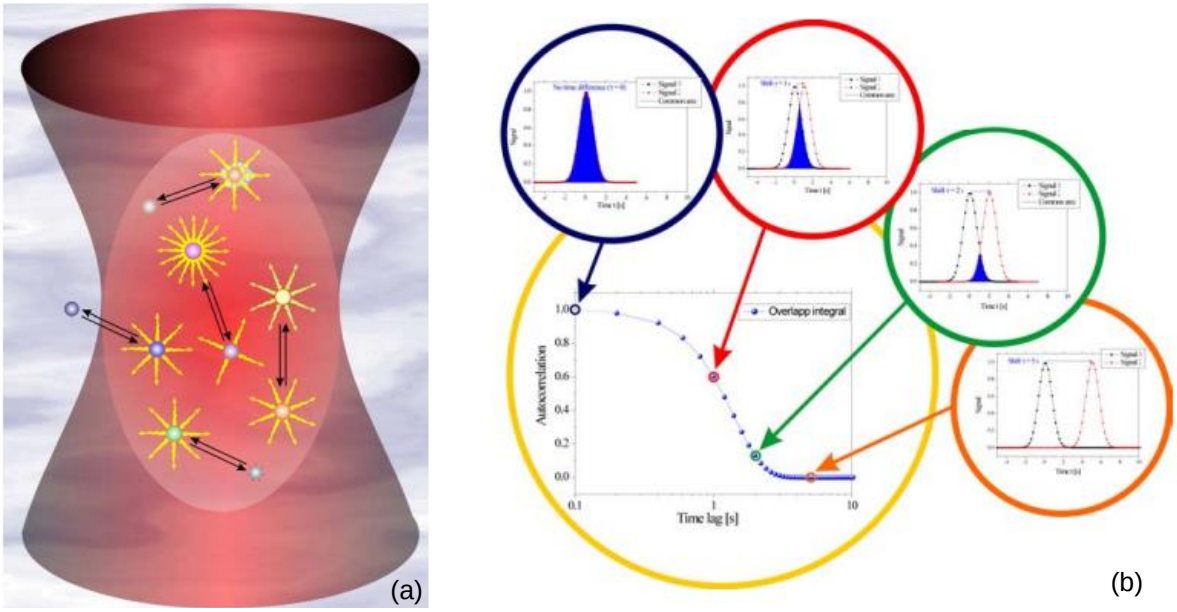


Figure 1. (a) Image of confocal volume using focused light beam and a confocal aperture (b) Visual representation of how an autocorrelation is created via overlap

Data to Results

Since the confocal volume also affects the data, it is necessary to account for the size and shape. In this work you are going to find the translational diffusion time (τ_D) and translational diffusion coefficient (D) using autocorrelation technique. It is known that the translational diffusion coefficient of a molecule is related to its size by the Einstein-Stoke relation, given as

$$D = \frac{k_B T}{6\pi\eta r_H}$$

Where D is the translational diffusion coefficient of the molecule, k_B is the Boltzmann constant given as $1.380649 * 10^{-23} J \cdot K^{-1}$, T is the temperature of the sample, and η is the viscosity of water given as $1.0016\ mPa \cdot s$ at 20 °C, and r_H is the hydrodynamic radius of the sample of interest.

The translational diffusion coefficient (D) is related to the correlation time (τ_D) as

$$\tau_D = \frac{\omega_{xy}^2}{4D}$$

Where ω_{xy} is the radius of confocal detection volume in the xy plane. In the following procedure, τ_D will experimentally be determined using Kristine and python fitting at varying concentrations for Rhodamine 110, a small fluorescent dye that is well classified with a translational diffusion coefficient of $4.4 * 10^{-10}\ m^2 \cdot s^{-1}$. This will allow the radius of confocal detection volume (ω_{xy}) to be determined. Thus, allowing the hydrodynamic radius (r_H) of GFP to be determined.

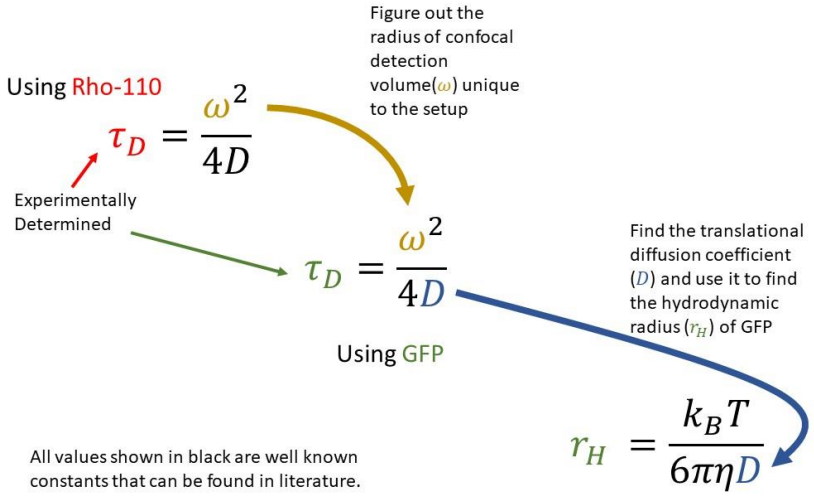


Figure 2. Flow chart explaining how a well-known system (Rho-110) can be used to determine an experimental setup parameter to gain information on an unknown system (GFP)

For those who are more interested

Similarly, the rotational diffusion correlation (D_θ) is related to the molecule's radius by

$$D_\theta = \frac{k_B T}{8\pi\eta r_\theta^3}$$

where r_θ is the radius of the probed molecule. Furthermore, the relation between rotational correlation time (τ_θ) and (D_θ) is given by

$$\tau_\theta = \frac{1}{6D_\theta}$$

where τ_θ will be determined experimentally using python script allowing $r_\theta$to be determined using equations (3) and (4) bypassing D_θ entirely, giving

$$r_\theta = \sqrt[3]{\frac{3k_B T\tau_\theta}{4\pi\eta}}$$

Data Fitting Theory

The autocorrelation function ($G(\tau)$) is defined as

$$G(\tau) = \frac{< F(t) \cdot F(t+\tau) >}{< F(t) >^2} \quad (1)$$

Where $F(t)$ is the fluorescence intensity at time t and τ is a delay time. One can image sliding the fluorescence intensity curve against its self and measuring the overlap for each time delay [see figure 1b]. One can use the definition of brightness and consider the volume and shape of the detection volume [see figure 1a] to obtain the following relationship [see references 2-3 for details]

$$G(\tau) = G(0)\left(1 + \frac{4D}{\omega_{xy}^2}\right)^{-1}\left(1 + \frac{4D\tau}{\omega_z^2}\right)^{-1/2} \quad (2)$$

Here ω_{xy} is the radius of the confocal detection volume in the xy-plane while ω_z is the radius in the z-direction. $G(0)$ is the amplitude of the correlation curve at $\tau = 0$ and D is the translational diffusion coefficient defined previously.

Equation 5 is a simplification of Fit #26 used by Kristine to determine the confocal detection volume. Kristine uses a slightly more complex fitting to account for various fluorescent processes as well as detector settings.

Including the triplet state the equation becomes

$$G(t) = b + \frac{1}{N}.\frac{1}{1+\frac{x}{|\Upsilon_d|}}.\frac{1}{\sqrt{1+\frac{x}{s^2.|\Upsilon_d|}}}\left(1 - |B_T| + |B_T|e^{\frac{-x}{|T_t|}} - |Ap| + |Ap|.e^{\frac{-x}{|T_p|}}\right) \quad (3)$$

Where *b=baseline, N=number of molecules in excitation volume,* τ_d*=diffusion time, s=geometric factor for excitation volume and treated as fixed, Bt = amplitude triplet state, Tt = triplet kinetic time, Ap = Amplitude photophysical, Tp = triplet/photophysical time .*

Also make sure we are using single exponential fit for RHO 110 and two exponentials for GFP due to their intrinsic nature *.*

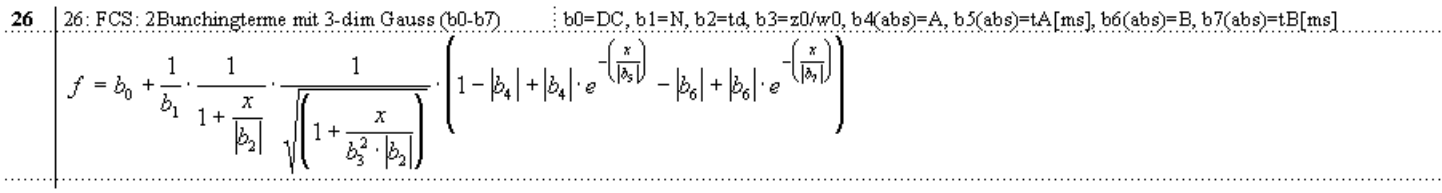


Figure 3. Fit #26 as used by Kristine. Important parameters are b2 corresponding to the translational diffusion time (τ_D).

References:

1 . Gel, C., Brockwell, D., Smith, A. e*t al. Handbook of single molecule fluorescence spectroscopy* (Oxford University Press on Demand, 2006).

2 . Lakowicz, J. R. *Principles of Fluorescence spectroscopy* (Springer science & business media, 2013).

3 . http://www.fcsxpert.com/classroom/theory/autocorrelation-diffusion-3d.html

Experiment

Material

1. Sample
    a. Rho-110 for standard curve
    b. Protein of interest Green Fluorescence Protein (GFP)
    c. Carbon filtered Buffer for blank pH=7.5
2. Carbon filtered deionized water (ddH2O) .
3. Glass slides

4. Pipette and tips
5. Lens paper (it is very important to only use lens paper when cleaning the objective)

Procedure

1. Prepare your samples in the wetlab, Kinard 217.
    a. For this experiment, you will need two samples: 4 nanomolar (nM) solutions of Rho-110 and a 10nM solution of GFP.
        i. You will be provided with a 100nM stock of Rho-110 and GFP.
        ii. Micropipette 96µL of buffer solution (carbon filtered ddH20 for Rho-110).
        iii. Add 4µL of 100nM stock of interest (Rho-110). This brings the volume to 100µL total.
        iv. Micropipette 98µL of buffer solution (0.1M PBS for GFP).
        v. Add 2µL of 100nM stock of interest (GFP). This brings the volume to 100µL total.

    b. You will also need to prepare carbon filtered ddH2O and PBS buffer solutions.
        i. This will be prepared and done for you. Just make sure to bring some to the confocal set up (Kinard 210).
2. Take your samples to the confocal setup in Kinard 210, for data collection and measurement.
    a. Glass slides, pipette and tips as well as lens paper will already be in Kinard 210.
3. The laser setup should already be turned on when you arrive. The parameters of the laser are controlled using Sepia II. **Please do not alter or change any of the laser parameters.**

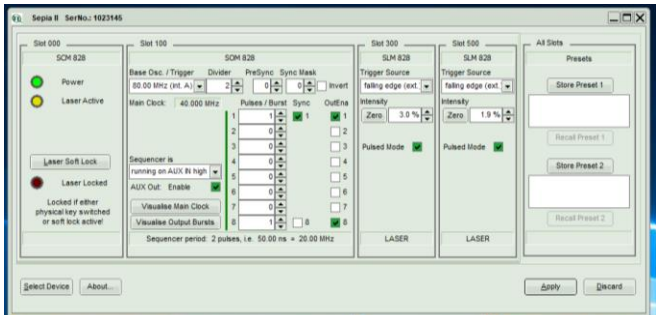


Figure 4. Screenshot of Sepia II, the laser control software. Do not alter these settings.

4. **Using lens paper**, clean the confocal microscope objective.

5. Pipette 50µL of carbon filtered ddH2O on top of the objective. The H2O will keep the index of refraction between the sample, the glass slide and the objective similar, allowing for better resolution.
6. Carefully place the glass slide on top of the ddH2O droplet on the objective.
7. Pipette 50µL of your sample (this should be your blank, PBS buffer) onto the glass slide directly above the objective for measurement. Use the light of the laser to ensure your sample drop is centered.
8. Focus the confocal microscope on your sample using the fine adjustment knob (the smaller rotator).
    a. First find the z-axis location where the laser light is a perfect pinpoint. This corresponds to the laser hitting the bottom surface of the glass slide. (Location S)
    b. Rotate the fine adjustment knob 1.5 turns counterclockwise until you find where the laser hits the top of the glass slide. (Location T)
    c. Rotate the fine adjustment knob 1.5 turns counterclockwise past the top to the glass slide to focus the objective in the center of your sample. (Location X)

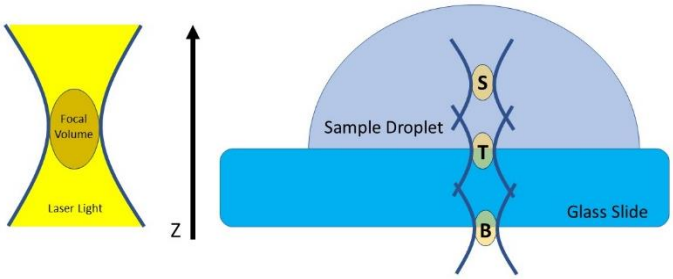
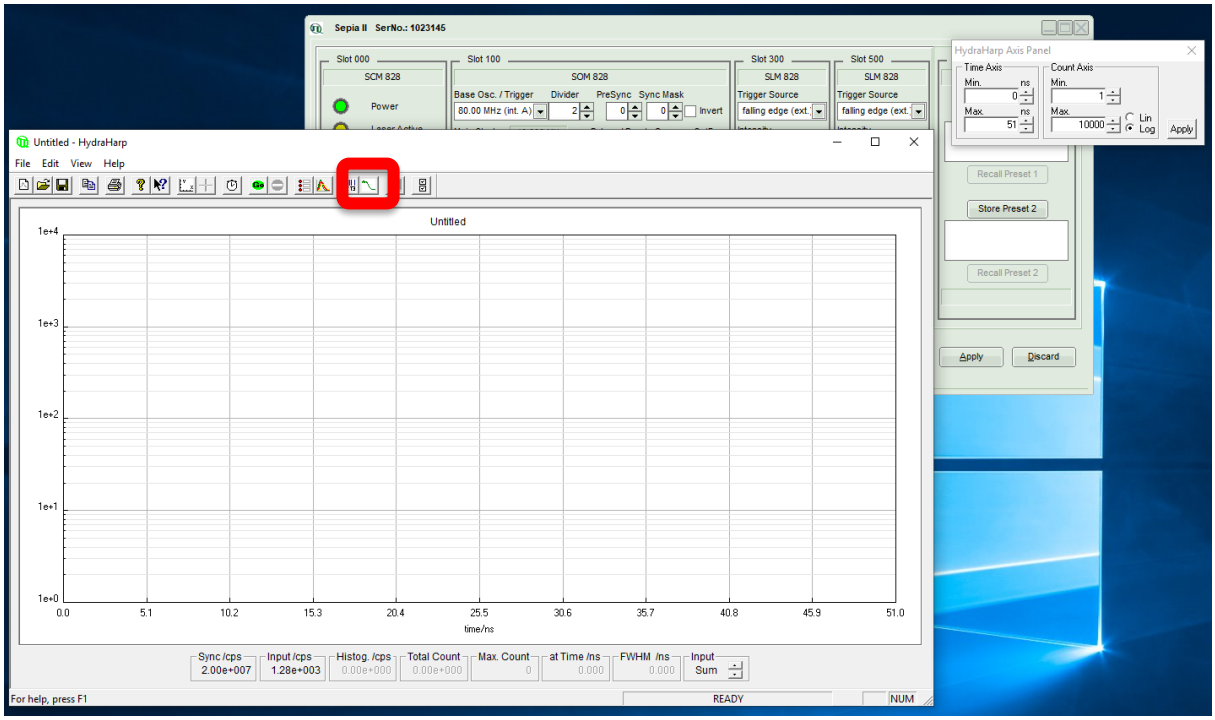


Figure 5. The microscope aperture focuses the laser light to create a focal volume where measurements will be detected. This focal volume remains in place while adjusting z-direction of the glass slide with the fine adjustment knob affects the focal volume location on the sample stage. At the boundary of the glass slide, the laser light is reflected, appearing as a sharp spot.

Data Collection

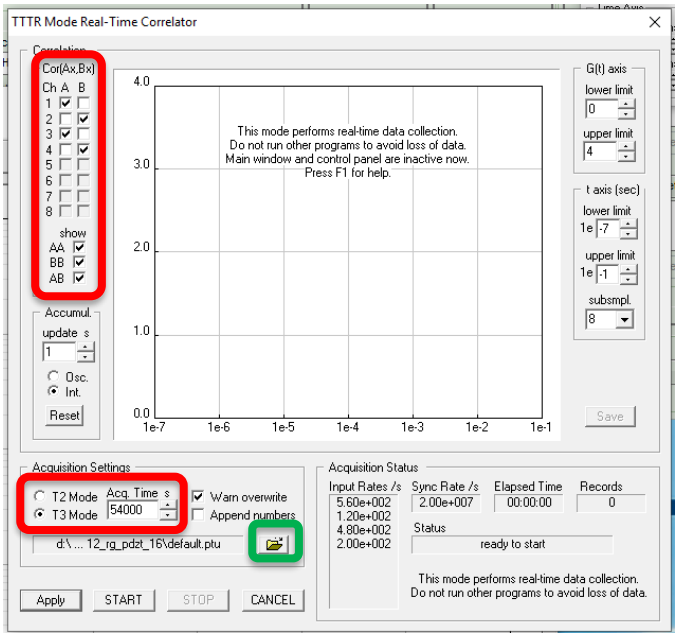
1. Open Hydraharp from the desktop by clicking on the icon.

2. Select TTTR mode (shown in the red box). A popup box for setting should appear.

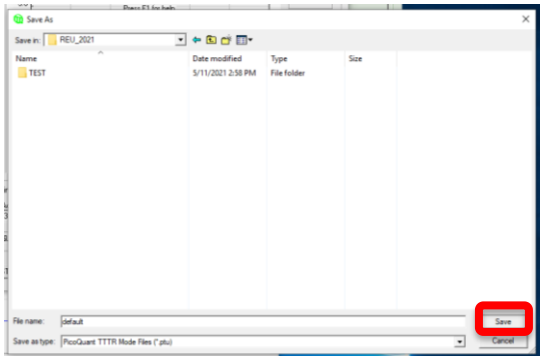


3. Select the appropriate measurement parameter (noted in the red boxes) and change the acquisition time for T3 mode (this may not be necessary but confirm the setting) to match that of your measurement. Note: Time is measured in seconds. If you want a 5 minutes measurement, you need to type 300.

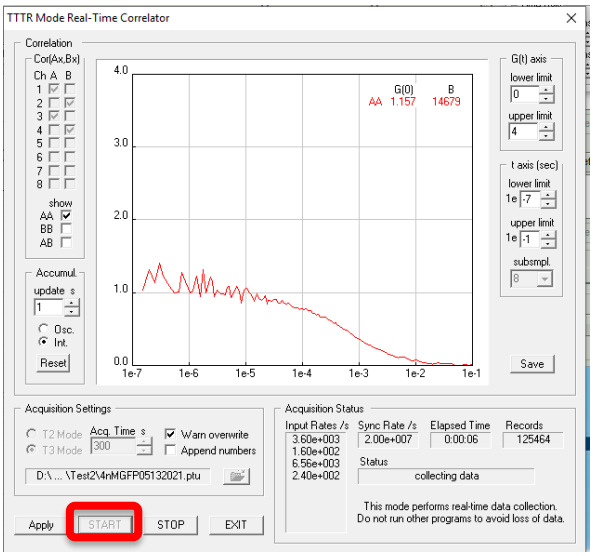
4. Then select the file location to save the measurements (see the green box).



5. Select an appropriate location to save your measurement files. Remember to include proper identifiers such as the date, your group name, what is the sample, and concentration. Suggestions: In the REU_2021 file, create a new file under your group's name. You can save all your measurements here. Click 'Save' to save your file location (shown via red box). The file explorer window should automatically close.



6. Then select 'START' on the TTTR Mode Real-Time Correlator window to start the measurement. A window like the one below should appear. (The image to the right shows the location of the start button in the red box. Note: It is gray in the image since a measurement is currently being taken.)

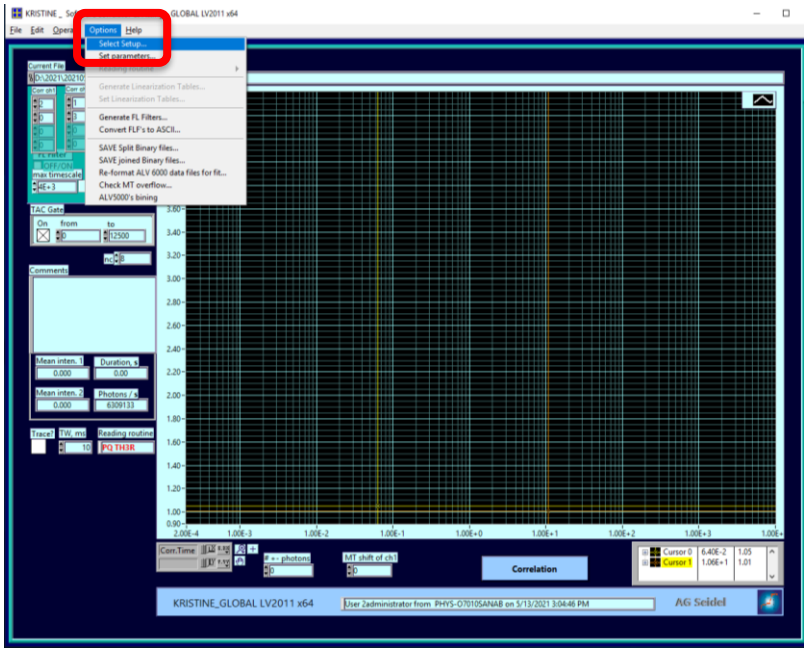


7. Repeat steps 7 and 8 in the Procedure followed by steps 4 and 5 of Data Collection to collect measurements for the following samples
    a. 4nM Rho-110
    b. 2nM Rho-110
    c. 10nM GFP
    d. 10nM GFP(Heated)

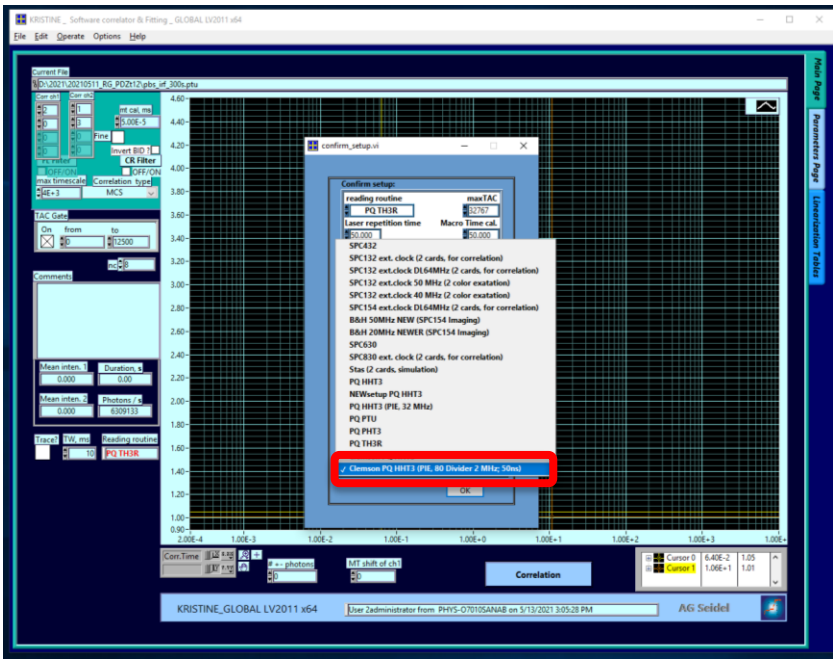
    Ask your TA for assistance with dilutions if you are unsure. It is recommended to mix the dilutions directly on the glass coverslip.

Data Analysis

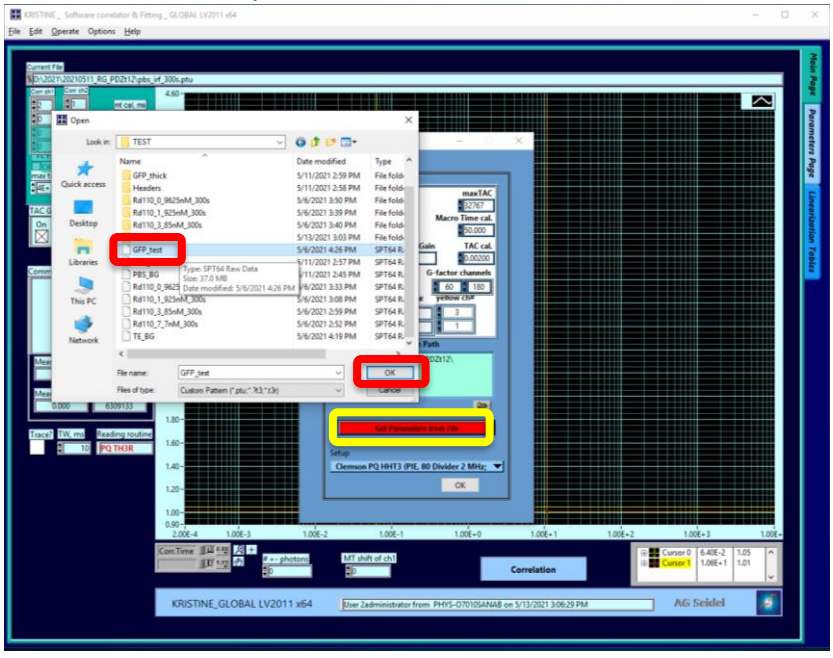
1. Using File Explorer, open Kristine for data analysis. Kristine will allow you to split your data files to lessen the workload of the computer and determine the diffusion lifetime for your molecule of interest.
2. Once Kristine is open, start by selecting the setup. (Options → Select Setup…).



3. This will prompt a setup window to appear. Select “Clemson PQ HHT3 (PIE, 80 Divider 2 MHz; 50ns)” (shown by the red box) to let Kristine know the parameter used for data collection. This ensures the appropriate fit.

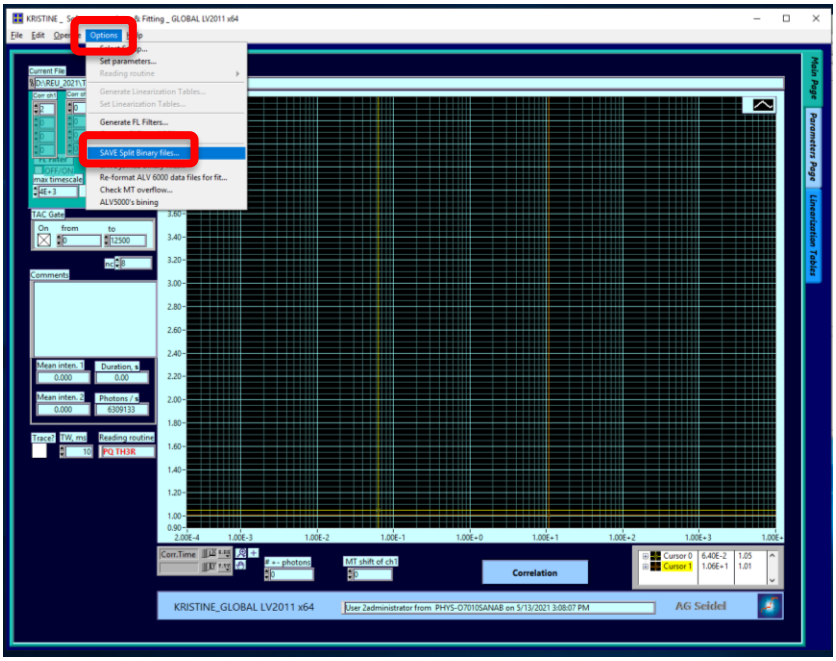


4. You will “Get Parameters from File” (the red button in the yellow box shown for emphasis). This prompts a file explorer window to appear. Select one of your data files for the computer to analyze the confocal setup parameters. Then click okay. (See the red boxes.)

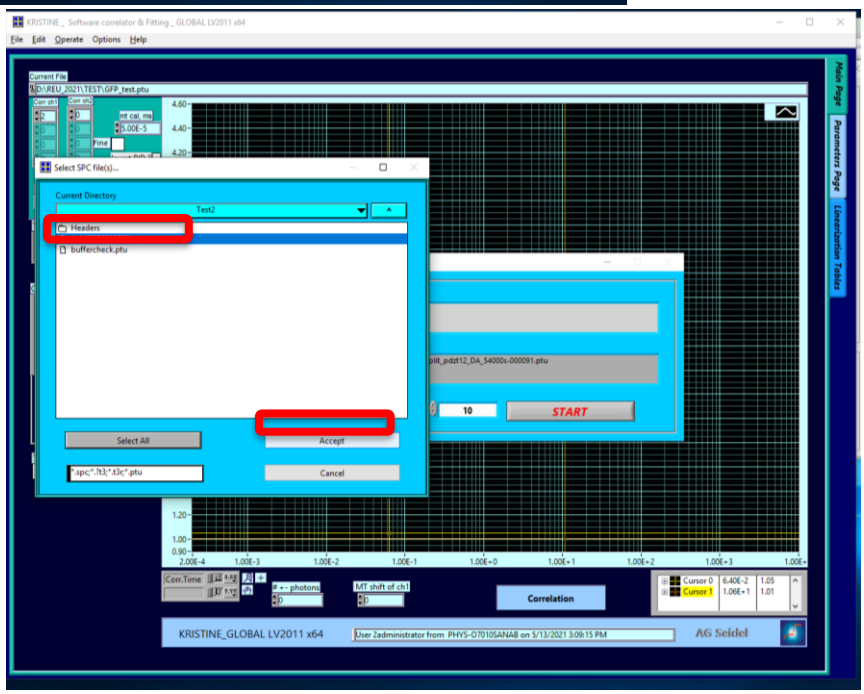


Splitting Files

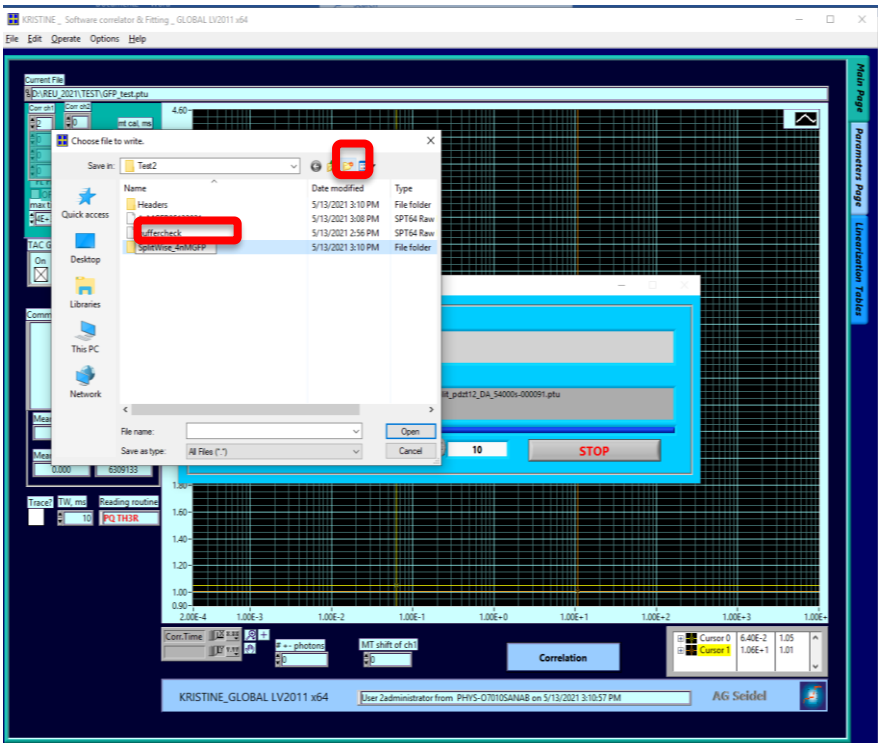
5. Your data files are currently too large for the desktop computer in 210 to handle. To solve this issue, we will split the one large file into smaller files by the number of photons. Go to Options → SAVE Split Binary Files (shown via red boxes).



6. Select the data file you want to split. Then click Accept. Make sure to only select one data file at a time, otherwise when the files are split, they will be combined into the same folder. This means if you have three data files, you will need to split files three times.



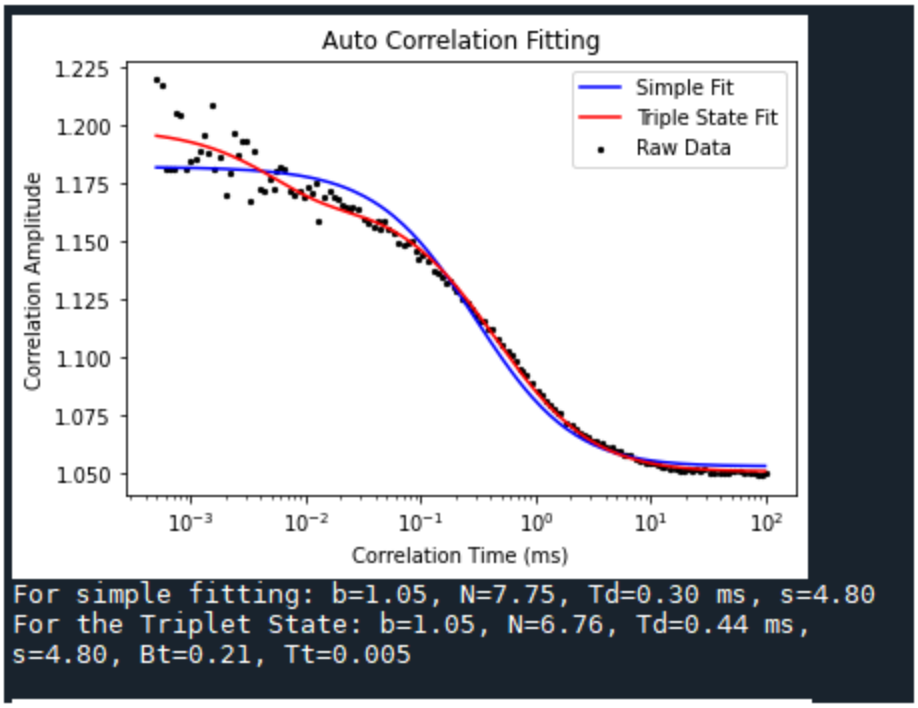
7. Make a new folder where the split files will be saved. This will make correlating all the files easier later. Select the folder. Make sure to name it appropriately.



8 Once the files are split you have to use the kristine again to save the file that can be use as input for analysis. For this just click Operate and from the pops up select correlate and once the file are loaded from the left top select file and save for fit , which will save you .COR file that can be analyzed by the python scripts.

Determining Diffusion Lifetime

You are provided with python script FCSfitting.py for Rho110 and FCSfitting_gfp.py for GFP which itself will determine the different parameters you are looking for with various fitted curve. Only the work you need to do is provide the path in the scripts where your files are. These values can be plugged in in the above equations as describe by the fig #2 . ie find ω from RHO 110 as Td and D is known for it , plug the value of ω to obtain rH the hydrodynamic radius for GFP. The other values required are obtained from the fit of GFP as show in figure below. Finally, you would compare the rH of GFP among various other methods and compare. Make sure in simulation work the obtained value is radius of gyration rG and is related to rH as rH =0.8* rG for spherical object.



Denaturation

Denaturation is process of breaking of many weak linkages or bond within a protein molecule that are responsible for highly ordered structure of the protein in its native state. Denaturation can be brought about in various ways—e.g., by heating, by treatment with alkali, acid, urea, or detergents, and by vigorous shaking depending on type of bond and interest of experiments. Here we are using heat as denaturing agent. On heating the barrel shape of GFP deforms and we are going to lose the fluorescent over time. This will be illustrated by heating the GFP at different temperature like 50°C 60°C ,70 °C ,80 °C, 90°C etc for 10 minutes and do the measurement and analysis again. The significant change is seen in terms of the amplitude of the fluorescent in autocorrelation curve as

$G(t) \approx \frac{1}{N}$ and N being the number of fluorescent particles on the confocal volume .The curve shown below illustrate the same molar GFP one being heated at 70°C with larger amplitude(white curve) and the other(greenish) native one.

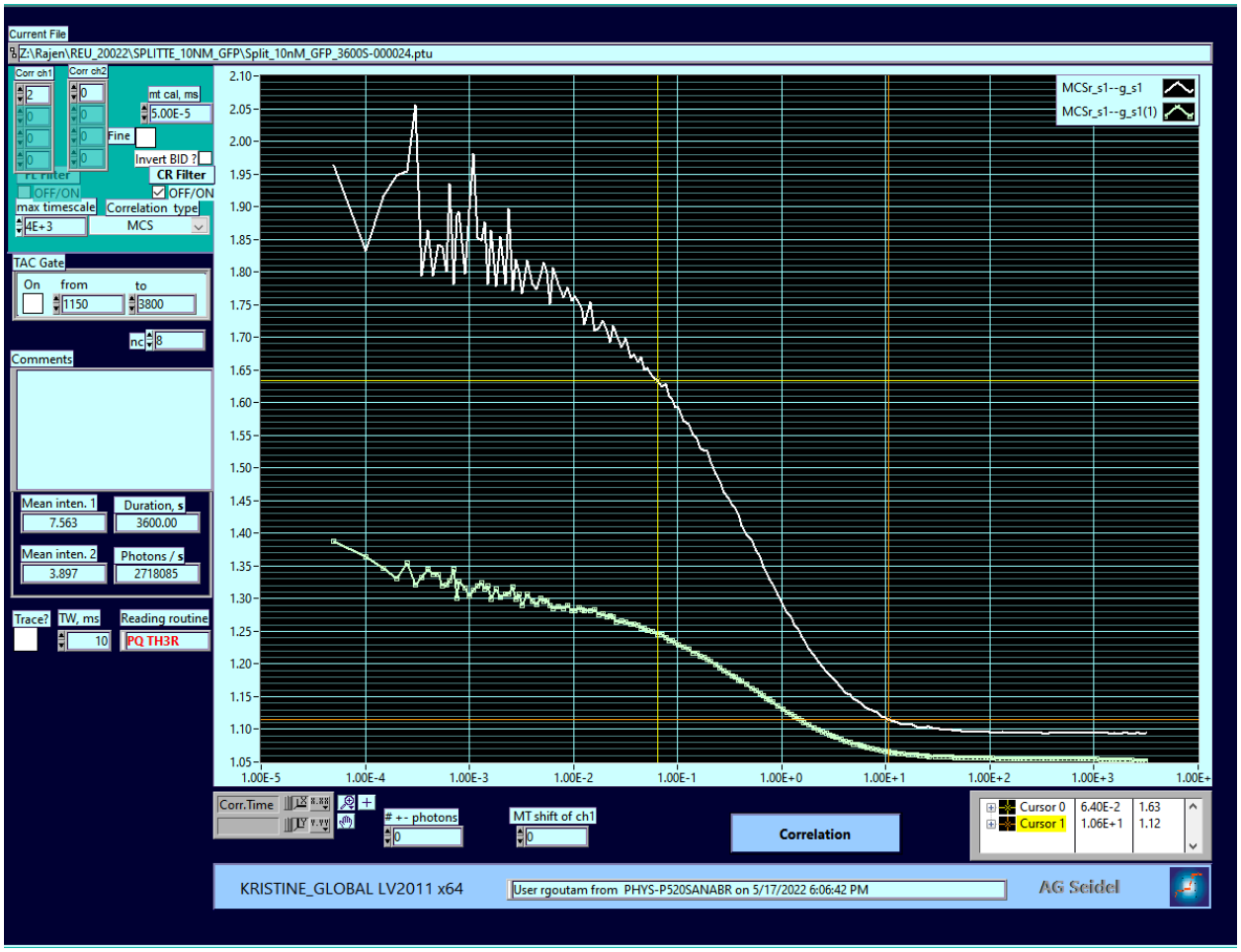
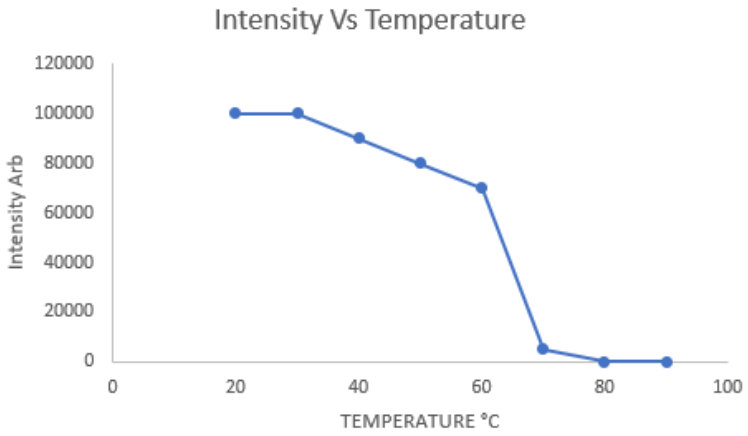


Fig illustration of denaturation of GFP by heating , Native state(greenish) has lower amplitude than the denatured(white)

Further If one we consider doing a long denaturation over a range of temperature the fluorescent property of GFP faded with increasing temperature and we will expect non fluorescence (only background) which is represented by the following diagram



Questions and Remarks (to be answered on Canvas)**:**

- Give a short summary (<200 words) regarding the methods used today.

- What are some benefits of using this method? What are some limitations? When would a researcher look to use this method?

- What data is measured in this method? What results are obtained from this method? How does the data and results relate? What is the range of error? How accurate are the results?

- Give a summary of the results you obtained today.

Appendix G

N2B Expression, Purification, Labeling Protocol

Protocols for N2B

Modifying Labeling

Danielle Latham

Aug 31, 2023

Expression:

For purposes of this protocol, I will assume the plasmid has already been transformed into the E. coli cell.

Media:

- LB Media for Overnight Growth
    - 10mL – I suggest making 100mL within 250mL flask with our small autoclave and keeping sterile.
- LB for Expression
    - 500mL within a 1L flask – This will have to be sterile using the large autoclave on the third floor of Jordan.

Steps:

1) Within a 50mL conical tube, place 10mL of LB media and add appropriate antibiotic (for the N2B samples this is ampicillin. If using 100mg/mL stock, add uL worth of ampicillin that you have mL of LB media.)
2) Add transformed cells (either from plate or glycerol stock) into 50mL conical tube with antibiotics and LB.
3) Angle the 50mL conical tube to maximize surface contact for oxygen intake. Incubate and shake at 37*C/225rpm overnight.
4) In the morning, measure the optical density (OD) and make a glycerol stock if desired.
    a. You do not have to make a glycerol stock if we already have many, but I like to keep at least two around in case one becomes contaminated. Glycerol stocks should be 10-20% glycerol evenly with the overnight cell culture. Since freezing and thawing affects cell viability, glycerol stocks should be kept small 50uL – 250uL.
5) Place 1L flask with LB media into the incubator to warm it up. Add appropriate antibiotic (for N2B this is 100ug/mL of ampicillin) and let the flask rock to mix.
6) Add the remaining overnight cell culture to the 1L flask. This is going to be your cell flask.
    a. To keep the solution sterile, do NOT add the cells which you measure OD for.
7) Measure the starting OD for the cell flask and store 1mL within a 1.5mL centrifuge tube to check expression later.
8) Record the OD of your flask every 30 minutes to 1 hour, storing a 1mL fraction for later use.

9) When the OD becomes close to 0.60, approximal 2 hours, add 1uM of IPTG (if using the 1M IPTG stock, add uL worth of IPTG that you have mL of LB media).
    a. IPTG can expire quickly when in solution. It should be made fresh (less than 2 months) and avoid multiple freezing and thawing cycles.
10) Record the OD of your flask every 30 minutes to 1 hour, storing a 1mL fraction for later use, until the OD becomes 1.2 or stabilizes for 1 hour.
11) Spin down the cell culture. You need to use the appropriate centrifuge tubes and the ultracentrifuge on the third floor of Jordan.
    a. I typically do 12,000 rpm for 10 minutes at 4*C.
    b. It helps with the yield if you chill the cells on ice before the spin down.
12) Remove the LB supernatant and freeze the cell pellet at -20*C.
    a. Typically, I will split the cell pellet into 1 gram fractions. This is done by determining the size of the pellet. Resuspending the pellet into 10mL of LB for each gram. Transferring the resuspended pellet into 50mL conical tube. Then spinning the pellet down again for freezing. This allows multiple attempts for purification as well as frees the centrifuge tubes for others use.
13) With your collected fractions: Spin them down into cell pellets. Remove the LB supernatant. Run an SDS page gel with the cell pellets to check for expression.
    a. Normalizing by OD helps, but is only necessary when troubleshooting.
    b. I typically add 80uL of 2x Sample Loading Buffer and pipette to mix before heating.

Purification:

Buffers:

- Lysis Buffer (10 mL per sample):
    - 8M Urea
    - 100mM NaH_2PO_4
    - 10mM TrisBase
    - pH 8.0 (w/ strips = Urea will destroy the probe)
    - 1mM BetaME (added fresh)
- Ni-NTA Buffers:
    - 20mM $NaPO_4$ pH 7.4
    - 300mM NaCl
    - 4M Urea
    - Imidazole varies:
        - Equilibration = 10mM
        - Wash = 25mM
        - Elution = 250mM
- Dialysis Labeling Buffer
    - 20mM $NaPO_4$ pH 7.4
    - 300mM NaCl
    - 4M Urea

- 1mM TCEP

Steps:

1) Resuspend cell pellet with 10 mL of Lysis Buffer
2) Sonicate resuspended cells with 60% amplitude, 1 second on – 2 seconds off, for 1 minute. Let the solution cool for five minutes, the repeat pervious sonication.
    a. Remember to always sonicate on ice to prevent sample overheating.
    b. Remember to always follow hearing safety and ensure the probe is submerged without touching the container.
    c. Make sure to collect sample of the resuspension (pre-sonication) and after lysis, but before spin down (post-sonication) for an SDS page gel.
3) Spin the sample down in the centrifuge 4,000*g for 30 minutes.
4) Collect the supernatant for histag purification using the Ni-NTA column.
    a. Freeze the pellet for later use or discard the pellet in biohazard waste
5) Follow HisPur Ni-NTA Spin Columns ThermoScientific User Guide for HisTag Purification
    a. Follow the denaturing condition protocol.
    b. Make sure to collect samples of the load, flow thru, and wash for SDS page gel to check purification.
    c. Spin columns can be used repetitively, remember to clean after use and note on the column how many times it has been used.
6) Use the nanodrop to record the concentration of the elution from the Ni-NTA columns.
7) Combine the elution fractions. Set up dialysis within the Dialysis Labeling Buffer overnight at 4*C.
    a. This is to remove the imidazole from the solution to rebind to the Ni-NTA beads for labeling.
    b. I typically used 1.5L for dialysis within the cylindrical glass containers with a lid.
    c. Ensure the dialysis solution is stirring to help drive the chemical gradient.
    d. You can place multiple samples within the same overnight dialysis solution.
8) The next morning, remove the sample from dialysis. Remeasure the concentration using the nanodrop.
9) Flash freeze extra solution with Liquid Nitrogen and stored at -80*C. Aim for 1.0 mg (or 1mL) aliquots.

Labeling with Histag

- Labeling Buffer:
    - 50mM NaH_2PO_4
    - 500mM NaCl
    - 2M Urea
    - 1mM TCEP
    - pH 7.5
- Ni-NTA Equilibration Buffer
    - 20mM $NaPO_4$ pH 7.4

- 300mM NaCl
- 4M Urea
- 10mM imidazole

1) Within a 1.5mL microcentrifuge tube, place 1mL of Ni-NTA Equilibration Buffer and 50uL of Ni-NTA Resin beads. Mix the beads via inverting 4-6 times. Gently, spin the beads down (1,000*g for 5 minutes). Remove the supernatant, leaving the beads in the microcentrifuge tube.
    a. This removes the storage buffer and preps the beads for binding.
2) Add 1mg of protein to the microcentrifuge tube with the beads.
    a. This should be a -80*C aliquot for the proteins - if it was prepared correctly. The aliquots vary in volume but are normalized in mass.
3) Allow the microcentrifuge tube to head over tail rock at 4*C for 2 hours.
    a. This allows the binding to the beads.
4) Gently, spin the beads down (1,000*g for 5 minutes). Remove the liquid, saving a sample to check binding via SDS gel.
    a. I also suggest checking this concentration on the nanodrop to determine the amount of protein that was bound.
5) Gently add 100uL of Labeling Buffer to the microcentrifuge tube. Wrap the microcentrifuge tubes with Aluminum to protect the dyes from light exposure.
6) Add 30% of cystine concentration (15% protein concentration) of Alexa488 maleimide and rock at room temperature for 2 hours.
    a. This is 0.64uL of 10mM Alexa488 maleimide.
    b. Make sure you are using the maleimide linkers or else the dyes will not bind to the cystines.
7) Wash off excess dye by gently adding 1.5mL of labeling buffer to wash the beads. Let the beads rock, head over tail, for 5 minutes at 4*C.
8) Spin the beads down (1,000*g for 5 minutes). Remove the liquid, leaving the beads behind.
9) Repeat steps 7-8 three times, or until the solution is clear after rocking.
    a. Keep checking beads for dye colors (aqua)
    b. Check for Donor (yellow) coloring on bead.
10) Resuspend the bead in 150uL of labeling buffer.
11) Remove 50L of evenly mixed solution from the microcentrifuge tube into another microcentrifuge. This will be your donor only (DO) fraction. Store in 4*C until the elution steps (step 17).
    a. You want to make sure to take some of the beads with you. Make sure to mix/flick the tube before removing the fraction, otherwise the beads might settle to the bottom.
12) Add 300% cystine concentration of Alexa647 maleimide to remaining sample and rock at 4*C overnight. This will be your double labeled (DA) fraction.
    a. This is 4.8uL of 10mM Alexa647N maleimide
13) With DA fractions, spin the beads down (1,000*g for 5 minutes).
14) With DA fractions, gently add 1.0mL of labeling buffer to wash the beads. Let the beads rock, head over tail, for 5 minutes at 4*C.
15) With DA fractions, spin the beads down (1,000*g for 5 minutes). Remove the liquid, leaving the beads behind.

16) Repeat steps 14-15 three times, or until the solution is clear after rocking.
    a. Keep checking beads for dye colors (aqua)
17) After the DO and DA fractions are washed, elute them from the Ni-NTA beads.
    a. This is done by adding 250mM imidazole to the labeling buffer. The elution should happen very quickly. No need to wait longer than 5 minutes.
    b. For DO: fill solution to 125uL with labeling buffer (This should mean adding 75uL to the 50uL already there) then simply add 16uL of 2M imidazole to the 125uL of labeling buffer.
    c. For DA: after removing the last of the wash buffer, resuspend in 375uL of labeling buffer and add 100uL of 2M imidazole
18) Gently, spin the beads down (1,000*g for 5 minutes). This time collect the solution.
    a. Here should be your eluted, labeled protein.
19) Use the confocal set up to determine the degree of labeling (how much DA labeling do we really have).
    a. See next section for details
    b. Nanodrop can also be used for degree of labeling.
20) Samples can be aliquoted into 15-25uL and flash frozen with Liquid Nitrogen and stored at -80*C.

Determining Degree of Labeling

1) Set up coverside.
2) Measure BG to ensure buffer is clean.
3) Add barley any sample for single molecule measurements.
    a. This may require a mid-dilution step
4) Use Paris to ensure single molecule and modify accordingly.
5) After stabilization, measure for 30 minutes
6) Check labeling with various software:
    a. Move data to server
    b. Kristine to split files
    c. Paris for burst selection
    d. Margarita to visualize

Varying Labeling methods

Method 1: Increasing Dye concentration

Determine which dye is lacking.

- If we need more donor:
    - Increase the donor concentration to 50%xC from 30%xC.

        - This means adding 1.07uL of 10mM Atto488
- If we need more acceptor:
    - Increase the acceptor concentration to 600%xC from 300%xC.
        - This means adding 9.6uL of 10mM Atto647N
    - Incubate at overnight at room temperature
        - This means incubating at room temperature for Labeling step 12

Method 2: Labeling both dyes at once

Simply add both dyes at the same time in varying ratios. Ensure the total dye concentration is always 300%xprotein.

- Instead of typical Labeling steps 6-12, you will add both Alexa488 and Alexa647 at the same time and incubate at 4*C overnight (I might decide to change this to RT later). This means the total dye volume will be 4.8uL if using the 10mM stocks.
    - 50% Alexa488 : 50% Alexa647
        - **Do this one first then see how it needs to be adjusted!**
        - 2.46uL of 10mM Alexa488
        - 2.46uL of 10mM Alexa647N
    - 30% Alexa488 : 70% Alexa647
        - 1.44uL of 10mM Alexa488
        - 3.36uL of 10mM Alexa647N
    - 40% Alexa488: 60% Alexa647N
        - 1.92uL of 10mM Alexa488
        - 2.88uL of 10mM Alexa647N
    - (I might add more options later)
- There will not be a DO fraction for this labeling. That is okay.
- Follow the wash and elution steps as usual

Method 3: Acceptor Pre-Label

Here we are doing to add a little Alexa647 (50%xC) for 30 minutes. Wash 3x. Then adding 50%xC Alexa488. Waiting two hours at room temp. Washing 3x, then adding 300%xC Alexa647 incubate overnight at 4*C. Wash 5x. There will be no DO fraction in this method either.

- After Labeling step 5, Add 1.07uL of 10mM Alexa647. Let it incubate, head over tail rocking, for 30 minutes.
- Wash the excess dyes 3x times.
- Continue with Labeling step 6, except add 1.07uL of Alexa488 (50%xC).
- Continue as normal, except there will be no removal of a donor only fraction (Labeling step 11).

12% SDS Gels:

- Expression Gel:
    - Cell OD fractions to ensure protein is expressing correctly
- Ni-NTA Purification Gel (Coomassie Blue):
    - Check sample is purified from Ni-NTA Column
        i. Ladder
        ii. Pre-Sonication
        iii. Post-Sonication
        iv. Ni-NTA Load
        v. Ni- NTA Flow Thru
        vi. Ni-NTA Wash
        vii. Ni-NTA Elution
- Labeling Gel:
    - Check to ensure the samples are labeling correctly
      Image for dyes before staining.
        i. Ladder
        ii. Post-Dialysis
        iii. Post-binding (Step 4)
        iv. Final Wash (D)
        v. Post-Labeling (DO)
        vi. Final Wash (A)
        vii. Post-Labeling (DA)

We can also do SEC for analytics

Appendix H

Western Blotting Protocol

Western Blotting

Prepare and run two similar SDS gels following typical protocol. One gel will be used for typically gel imaging (Coomassie Blue). The second gel will be used for membrane transfer. Gels that have been stained with Coomassie Blue will not transfer to the membrane.

Immediately after SDS gels finishing running, start protocol for protein transfer to membrane

NOTES:

The membrane should never be touch with gloves. Always use forceps when moving the membrane.

The protocol is for PVDF membrane with 6xHisTag Alexa488 antibody, but can be modified for other membranes and antibodies.

Membrane Transfer

This protocol uses the Trans-Blot Turbo and Trans-Blot Turbo RTA Transfer kit, PVDF (Cat. #1704272) from BioRad following the given protocol.

1) Mix transfer buffer – the following is for a single transfer (1 sample) multiply the recipe to match your number of samples
    a. 16 mL (1 part) 5xTransfer Buffer
    b. 48 mL (3 parts) ddH2O
    c. 16 mL (1 part) ethanol
        i. The solution will turn cloudy upon the addition of ethanol
2) Wet Transfer Stacks and Membrane
    a. PVDF membrane
        i. Soak in 100% methanol until transparent
            - Membrane will not become completely clear, but opaque
        ii. Soak in 30 mL of 1x Transfer buffer for 2-3 minutes
    b. Transfer Stacks
        i. Soak 2 transfer stacks separated via the blue divider
        ii. 50 mL of 1x Transfer buffer for 2-3 minutes
            - I place the transfer stacks in the plastic box first then pour the transfer solution on top to ensure good absorbance
            - Not all solution will be absorbed
3) Assembly of Cassette for transfer
    a. Taking the cassette from the Trans-Blot Turbo, open and stack
        i. 1st soaked transfer stack
        ii. Wetted membrane
        iii. SDS gel used for transfer
        iv. 2nd soaked transfer stack

b. Ensure all the parts are laid flat with no air bubbles. Use the roller to remove air bubbles and excess liquid.
c. Close and lock the cassette. Turn the cassette side ways to drain off excess liquid. This will help with the transfer.

4) Run the Transfer
    a. Place the cassette into the Trans-Blot Turbo and select the appropriate running setting
        i. I typically use the preset Turbo options and have gotten good results.
5) When completed –
    a. The membrane can be immediately blocked or dried for storage
        i. Membrane Storage
            - 4*C for
            - -20*C for
        ii. If blocking, ensure the membrane does not dry out. If the membrane does dry out, rehydrate with methanol.
    b. The SDS gel can be disposed or Coomassie stained to check the transfer

Blotting

Blotting involves using a mixture of unspecific proteins (typically nonfat milk) to cover the empty places on the membrane where proteins were not placed during the transfer. Membranes must be blotted after transfer to prevent antibodies from sticking to the empty membrane.

<u>Buffers:</u>

- TBST
    - 137 mM NaCl
    - 2.7 mM KCl
    - 19 mM TrisBase
    - pH 7.4
    - Filter 0.22 um
    - Add 1mL of Tween20 for each 20L of solution
        - It is important to add Tween20 **after** the solution is at the proper pH and filtered
        - Tween20 is highly viscous, slow, and careful while pipetting
- Milk Blocking Buffer
    - 5% milk into TBST = Add 5g of nonfat dry milk into 100 mL of TBST
    - Ensure mixture is well mixed and milk is fully dissolved
    - Gravity filter with filter paper

<u>Procedure:</u>

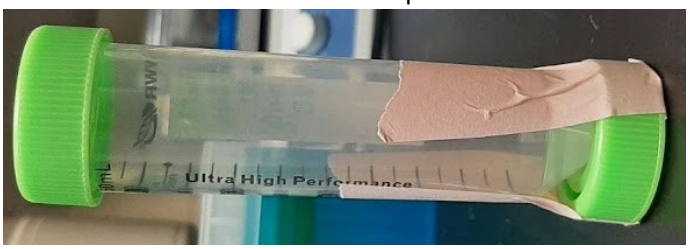
- Ensure membrane is hydrated.
    - If the membrane is dried, rehydrate with methanol then rinse with ddH2O
- Place membrane in enough blocking buffer to completely cover.
    - I use one of the glass boxes with 20mL of blocking buffer.
- Place membrane in blocking buffer under agitation (shaking) for 1 hour at 4*C

- Rinse membrane with TBST by dumping off blocking buffer, adding enough TBST to cover the swishing for 10 seconds.

Immunostaining: with conjugated, fluorophore antibody

This procedure assumes the antibody is conjugated with a fluorophore as well as already diluted and prepared. If you have questions on antibody dilution or preparation, see the instructions or notes from the antibody provider. Typically, our lab uses conjugated antibodies (the primary/tagging antibody comes with the marker/dye already attached). Most antibodies can be saved and used more than once. Some antibodies are light sensitive, thus the membrane and antibodies should be kept cover.

- The blocked membrane should be placed in a fresh solution of TBST containing the conjugated antibodies.
    - This can be done using 20mL of TBST antibody solution within glass box **OR**
    - Place membrane in a 50mL falcon tube ensuring the membrane does not overlap with itself. Add 3-5mL of TBST antibody solution
- Allow membrane with antibody solution to remain agitated (shaking) overnight at 4*C
    - If using a 50mL falcon tube, use the rocker. Type an extra lid to the bottom of the tube creating a perfect cylinder so that the tube rocks straight. Place the tube in the 'Antibody Rocking Box' (a cardboard box) to ensure the tube does not fall off the rocker overnight. A pipette box can be inserted into the rocking box if a small number of tubes are being used to ensure the tube does not end up stuck.



- The next morning, remove and save the antibody solution
- Add the same volume of TBST solution to the membrane. Agitate for 20 minutes at 4*C
- Drain off TBST, then add fresh TBST. Agitate for 20 minutes at 4*C
- The membrane is now ready for imaging with the ChemiDoc

REFERENCES.

1. Maruyama, K. 1997. Connectin/titin, giant elastic protein of muscle. *The FASEB Journal*. 11(5):341-345, doi: https://doi.org/10.1096/fasebj.11.5.9141500, https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.11.5.9141500.

2. Labeit, S., B. Kolmerer, and W. A. Linke. 1997. The Giant Protein Titin. *Circulation Research*. 80(2):290-294, doi: doi:10.1161/01.RES.80.2.290, https://www.ahajournals.org/doi/abs/10.1161/01.RES.80.2.290.

3. Levinthal, C. (1969). How to fold graciously.

4. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The Energy Landscapes and Motions of Proteins. *Science*. 254(5038):1598-1603, doi: doi:10.1126/science.1749933, https://www.science.org/doi/abs/10.1126/science.1749933.

5. Onuchic, J. N., and P. G. Wolynes. 1993. Energy landscapes, glass transitions, and chemical reaction dynamics in biomolecular or solvent environment. *The Journal of Chemical Physics*. 98(3):2218-2224, doi: 10.1063/1.464201, https://doi.org/10.1063/1.464201.

6. Saven, J. G. 2001. Designing Protein Energy Landscapes. *Chemical Reviews*. 101(10):3113-3130, doi: 10.1021/cr000058w, https://doi.org/10.1021/cr000058w.

7. Anfinsen, C. B. 1973. Principles that Govern the Folding of Protein Chains. *Science*. 181(4096):223-230, doi: doi:10.1126/science.181.4096.223, https://www.science.org/doi/abs/10.1126/science.181.4096.223.

8. Sela, M., F. H. White, and C. B. Anfinsen. 1957. Reductive Cleavage of Disulfide Bridges in Ribonuclease. *Science*. 125(3250):691-692, doi: doi:10.1126/science.125.3250.691, https://www.science.org/doi/abs/10.1126/science.125.3250.691.

9. Betancourt, M. R., and J. N. Onuchic. 1995. Kinetics of proteinlike models: The energy landscape factors that determine folding. *The Journal of Chemical Physics*. 103(2):773-787, doi: 10.1063/1.470109, https://doi.org/10.1063/1.470109.

10. Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*. 21(3):167-195, doi: https://doi.org/10.1002/prot.340210302, https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340210302.

11. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. 293(2):321-331, doi: 10.1006/jmbi.1999.3110, https://dx.doi.org/10.1006/jmbi.1999.3110.

12. Uversky, V. N. 2014. Introduction to Intrinsically Disordered Proteins (IDPs). *Chemical Reviews*. 114(13):6557-6560, doi: 10.1021/cr500288y, https://doi.org/10.1021/cr500288y.

13. Tsang, B., I. Pritišanac, S. W. Scherer, A. M. Moses, and J. D. Forman-Kay. 2020. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell*. 183(7):1742-1756, doi: 10.1016/j.cell.2020.11.050, https://dx.doi.org/10.1016/j.cell.2020.11.050.

14. Romero, P., Z. Obradovic, and A. K. Dunker. 2004. Natively Disordered Proteins. *Applied Bioinformatics*. 3(2):105-113, doi: 10.2165/00822942-200403020-00005, https://doi.org/10.2165/00822942-200403020-00005.

15. Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. Sequence complexity of disordered protein. *Proteins*. 42(1):38-48, doi: 10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3.

16. Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*. 16(1):18-29, doi: 10.1038/nrm3920, https://dx.doi.org/10.1038/nrm3920.

17. Pederson, T. 2011. The nucleolus. *Cold Spring Harb Perspect Biol*. 3(3), doi: 10.1101/cshperspect.a000638.

18. Brangwynne, C. P., C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, and A. A. Hyman. 2009. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*. 324(5935):1729-1732, doi: 10.1126/science.1172046, https://dx.doi.org/10.1126/science.1172046.

19. Phair, R. D., and T. Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature*. 404(6778):604-609, doi: 10.1038/35007077, https://dx.doi.org/10.1038/35007077.

20. Shin, Y., and C. P. Brangwynne. 2017. Liquid phase condensation in cell physiology and disease. *Science*. 357(6357):eaaf4382, doi: 10.1126/science.aaf4382, https://dx.doi.org/10.1126/science.aaf4382.

21. Alberti, S., A. Gladfelter, and T. Mittag. 2019. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*. 176(3):419-434, doi: 10.1016/j.cell.2018.12.035, https://dx.doi.org/10.1016/j.cell.2018.12.035.

22. Brangwynne, C. P., T. J. Mitchison, and A. A. Hyman. 2011. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences*. 108(11):4334-4339, doi: 10.1073/pnas.1017150108, https://dx.doi.org/10.1073/pnas.1017150108.

23. Banani, S. F., H. O. Lee, A. A. Hyman, and M. K. Rosen. 2017. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*. 18(5):285-298, doi: 10.1038/nrm.2017.7, https://dx.doi.org/10.1038/nrm.2017.7.

24. Wei, M.-T., S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne. 2017. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry*. 9(11):1118-1125, doi: 10.1038/nchem.2803, https://dx.doi.org/10.1038/nchem.2803.

25. Morrical, S. W., and M. M. Cox. 1990. Stabilization of recA protein-ssDNA complexes by the single-stranded DNA binding protein of Escherichia coli. *Biochemistry*. 29(3):837-843, doi: 10.1021/bi00455a034, https://dx.doi.org/10.1021/bi00455a034.

26. Chowdhury, A., A. Borgia, S. Ghosh, A. Sottini, S. Mitra, R. S. Eapen, M. B. Borgia, T. Yang, N. Galvanetto, M. T. Ivanović, P. Łukijańczuk, R. Zhu, D. Nettels, A. Kundagrami, and B. Schuler. 2023. Driving forces of the complex formation between highly charged disordered proteins. *Proc Natl Acad Sci U S A*. 120(41):e2304036120, doi: 10.1073/pnas.2304036120.

27. Laflamme, G., and K. Mekhail. 2020. Biomolecular condensates as arbiters of biochemical reactions inside the nucleus. *Communications Biology*. 3(1), doi: 10.1038/s42003-020-01517-9, https://dx.doi.org/10.1038/s42003-020-01517-9.

28. Pitchiaya, S., M. D. A. Mourao, A. P. Jalihal, L. Xiao, X. Jiang, A. M. Chinnaiyan, S. Schnell, and N. G. Walter. 2019. Dynamic Recruitment of Single RNAs to Processing Bodies Depends on RNA Functionality. *Molecular Cell*. 74(3):521-533.e526, doi: 10.1016/j.molcel.2019.03.001, https://dx.doi.org/10.1016/j.molcel.2019.03.001.

29. Nott, T., E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, David, T. Pawson, Julie, and Andrew. 2015. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*. 57(5):936-947, doi: 10.1016/j.molcel.2015.01.013, https://dx.doi.org/10.1016/j.molcel.2015.01.013.

30. Zaffagnini, G., A. Savova, A. Danieli, J. Romanov, S. Tremel, M. Ebner, T. Peterbauer, M. Sztacho, R. Trapannone, A. K. Tarafder, C. Sachse, and S. Martens. 2018. p62 filaments capture and present ubiquitinated cargos for autophagy. *Embo j*. 37(5), doi: 10.15252/embj.201798308.

31. Wilfling, F., C. W. Lee, P. S. Erdmann, Y. Zheng, D. Sherpa, S. Jentsch, B. Pfander, B. A. Schulman, and W. Baumeister. 2020. A Selective Autophagy Pathway for Phase-Separated Endocytic Protein Deposits. *Mol Cell*. 80(5):764-778.e767, doi: 10.1016/j.molcel.2020.10.030.

32. Yamasaki, A., J. M. Alam, D. Noshiro, E. Hirata, Y. Fujioka, K. Suzuki, Y. Ohsumi, and N. N. Noda. 2020. Liquidity Is a Critical Determinant for Selective Autophagy of Protein Condensates. *Mol Cell*. 77(6):1163-1175.e1169, doi: 10.1016/j.molcel.2019.12.026.

33. Zhang, G., Z. Wang, Z. Du, and H. Zhang. 2018. mTOR Regulates Phase Separation of PGL Granules to Modulate Their Autophagic Degradation. *Cell*. 174(6):1492-1506.e1422, doi: 10.1016/j.cell.2018.08.006.

34. Furqan, D., R. C. Samuel, M. M. Diana, H. P. Aaron, N. Gergely, C. L. Wellington, B. S. Christopher, C. Jeong-Mo, W. K. Richard, and V. P. Rohit. 2023. Biomolecular condensates form spatially inhomogeneous network fluids. *bioRxiv*.2023.2010.2007.561338, doi: 10.1101/2023.10.07.561338, http://biorxiv.org/content/early/2023/10/09/2023.10.07.561338.abstract.

35. Murthy, A. C., G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi. 2019. Molecular interactions underlying liquid−liquid phase separation of the FUS low-complexity domain. *Nature Structural & Molecular Biology*. 26(7):637-648, doi: 10.1038/s41594-019-0250-x, https://dx.doi.org/10.1038/s41594-019-0250-x.

36. Brangwynne, C., P. Tompa, and R. Pappu. 2015. Polymer physics of intracellular phase transitions. *Nature Physics*. 11(11):899-904, doi: 10.1038/nphys3532, https://dx.doi.org/10.1038/nphys3532.

37. Brady, J. P., P. J. Farber, A. Sekhar, Y.-H. Lin, R. Huang, A. Bah, T. J. Nott, H. S. Chan, A. J. Baldwin, J. D. Forman-Kay, and L. E. Kay. 2017. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences*. 114(39):E8194-E8203, doi: 10.1073/pnas.1706197114, https://dx.doi.org/10.1073/pnas.1706197114.

38. Niaki, A. G., J. Sarkar, X. Cai, K. Rhine, V. Vidaurre, B. Guy, M. Hurst, J. C. Lee, H. R. Koh, L. Guo, C. M. Fare, J. Shorter, and S. Myong. 2019. Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations. *Molecular Cell*. doi: 10.1016/j.molcel.2019.09.022, https://dx.doi.org/10.1016/j.molcel.2019.09.022.

39. Galvanetto, N., M. T. Ivanović, A. Chowdhury, A. Sottini, M. F. Nüesch, D. Nettels, R. B. Best, and B. Schuler. 2023. Extreme dynamics in a biomolecular condensate. *Nature*. doi: 10.1038/s41586-023-06329-5, https://doi.org/10.1038/s41586-023-06329-5.

40. Malinovska, L., S. Kroschwald, and S. Alberti. 2013. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta*. 1834(5):918-931, doi: 10.1016/j.bbapap.2013.01.003.

41. Protter, D. S. W., B. S. Rao, B. Van Treeck, Y. Lin, L. Mizoue, M. K. Rosen, and R. Parker. 2018. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep*. 22(6):1401-1412, doi: 10.1016/j.celrep.2018.01.036.

42. Alberti, S., S. Saha, J. B. Woodruff, T. M. Franzmann, J. Wang, and A. A. Hyman. 2018. A User's Guide for Phase Separation Assays with Purified Proteins. *Journal of Molecular Biology*. 430(23):4806-4820, doi: 10.1016/j.jmb.2018.06.038, https://dx.doi.org/10.1016/j.jmb.2018.06.038.

43. Zhou, H. X., V. Nguemaha, K. Mazarakos, and S. Qin. 2018. Why Do Disordered and Structured Proteins Behave Differently in Phase Separation? *Trends Biochem Sci*. 43(7):499-516, doi: 10.1016/j.tibs.2018.03.007, https://www.ncbi.nlm.nih.gov/pubmed/29716768.

44. Harmon, T. S., A. S. Holehouse, M. K. Rosen, and R. V. Pappu. 2017. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife*. 6, doi: 10.7554/elife.30294, https://dx.doi.org/10.7554/eLife.30294.

45. Batlle, C., I. Calvo, V. Iglesias, C. J. Lynch, M. Gil-Garcia, M. Serrano, and S. Ventura. 2021. MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. *Communications Biology*. 4(1):414, doi: 10.1038/s42003-021-01930-8, https://doi.org/10.1038/s42003-021-01930-8.

46. Banjade, S., Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu, and M. K. Rosen. 2015. Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proceedings of the National Academy of Sciences*. 112(47):E6426-E6435, doi: 10.1073/pnas.1508778112, https://dx.doi.org/10.1073/pnas.1508778112.

47. Harmon, T. S., A. S. Holehouse, and R. V. Pappu. 2018. Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New Journal of Physics*. 20(4):045002, doi: 10.1088/1367-2630/aab8d9, https://dx.doi.org/10.1088/1367-2630/aab8d9.

48. Li, P., S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. T. Nixon, and M. K. Rosen. 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. 483(7389):336-340, doi: 10.1038/nature10879, https://dx.doi.org/10.1038/nature10879.

49. Banani, S. F., A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen. 2016. Compositional Control of Phase-Separated Cellular Bodies. *Cell*. 166(3):651-663, doi: 10.1016/j.cell.2016.06.010, https://dx.doi.org/10.1016/j.cell.2016.06.010.

50. Rajshekar, S., O. Adame-Arana, G. Bajpai, K. Lin, S. Colmenares, S. Safran, and G. H. Karpen. 2023. Affinity hierarchies and amphiphilic proteins underlie the co-

assembly of nucleolar and heterochromatin condensates. *bioRxiv*. doi: 10.1101/2023.07.06.547894.

51. Weber, S., and C. Brangwynne. 2012. Getting RNA and Protein in Phase. *Cell*. 149(6):1188-1191, doi: 10.1016/j.cell.2012.05.022, https://dx.doi.org/10.1016/j.cell.2012.05.022.

52. Martin, E. W., and A. S. Holehouse. 2020. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg Top Life Sci*. 4(3):307-329, doi: 10.1042/etls20190164.

53. Borcherds, W. M., I. Alshareedah, A. Singh, T. Das, A. Bremer, P. R. Banerjee, R. V. Pappu, and T. Mittag. 2023. A stickers-and-spacers framework explains the viscoelastic and aging properties of biomolecular condensates. *Biophysical Journal*. 122(3, Supplement 1):296a, doi: https://doi.org/10.1016/j.bpj.2022.11.1674, https://www.sciencedirect.com/science/article/pii/S0006349522025905.

54. Yang, Y., H. B. Jones, T. P. Dao, and C. A. Castañeda. 2019. Single Amino Acid Substitutions in Stickers, but Not Spacers, Substantially Alter UBQLN2 Phase Transitions and Dense Phase Material Properties. *The Journal of Physical Chemistry B*. 123(17):3618-3629, doi: 10.1021/acs.jpcb.9b01024, https://doi.org/10.1021/acs.jpcb.9b01024.

55. Vernon, R. M., P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, and J. D. Forman-Kay. 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife*. 7:e31486, doi: 10.7554/eLife.31486, https://doi.org/10.7554/eLife.31486.

56. Joseph, J. A., A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara. 2021. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science*. 1(11):732-743, doi: 10.1038/s43588-021-00155-3, https://doi.org/10.1038/s43588-021-00155-3.

57. Babinchak, W. M., and W. K. Surewicz. 2020. Liquid–Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. *Journal of Molecular Biology*. 432(7):1910-1925, doi: 10.1016/j.jmb.2020.03.004, https://dx.doi.org/10.1016/j.jmb.2020.03.004.

58. Molliex, A., J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*. 163(1):123-133, doi: 10.1016/j.cell.2015.09.015.

59. Jawerth, L., E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, J. Mahamid, A. A. Hyman, and F. Jülicher. 2020. Protein condensates as aging Maxwell fluids. *Science*. 370(6522):1317-1323, doi: 10.1126/science.aaw4951.

60. Brehme, M., and C. Voisine. 2016. Model systems of protein-misfolding diseases reveal chaperone modifiers of proteotoxicity. *Dis Model Mech*. 9(8):823-838, doi: 10.1242/dmm.024703.

61. Romero, Obradovic, and K. Dunker. 1997. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Inform Ser Workshop Genome Inform*. 8:110-124.

62. Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*. 27(3):379-423, doi: 10.1002/j.1538-7305.1948.tb01338.x.

63. Li, X., P. Romero, M. Rani, A. K. Dunker, and Z. Obradovic. 1999. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform*. 10:30-40.

64. Obradovic, Z., K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, and A. K. Dunker. 2003. Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics*. 53(S6):566-572, doi: https://doi.org/10.1002/prot.10532, https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.10532.

65. Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. *Proteins: Structure, Function, and Bioinformatics*. 52(4):573-584, doi: https://doi.org/10.1002/prot.10437, https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.10437.

66. Linding, R., L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell. 2003. Protein Disorder Prediction: Implications for Structural Proteomics. *Structure*. 11(11):1453-1459, doi: https://doi.org/10.1016/j.str.2003.10.002, https://www.sciencedirect.com/science/article/pii/S0969212603002351.

67. Garner, E., P. Romero, A. K. Dunker, C. Brown, and Z. Obradovic. 1999. Predicting Binding Regions within Disordered Proteins. *Genome Inform Ser Workshop Genome Inform*. 10:41-50.

68. Radivojac, P., Z. Obradović, C. J. Brown, and A. K. Dunker. 2003. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*.216-227.

69. PENG, K., S. VUCETIC, P. RADIVOJAC, C. J. BROWN, A. K. DUNKER, and Z. OBRADOVIC. 2005. OPTIMIZING LONG INTRINSIC DISORDER PREDICTORS WITH PROTEIN EVOLUTIONARY INFORMATION. *Journal of Bioinformatics and Computational Biology*. 03(01):35-60, doi: 10.1142/s0219720005000886, https://www.worldscientific.com/doi/abs/10.1142/S0219720005000886.

70. Xue, B., R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky. 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 1804(4):996-1010, doi: https://doi.org/10.1016/j.bbapap.2010.01.011, https://www.sciencedirect.com/science/article/pii/S1570963910000130.

71. Peng, K., P. Radivojac, S. Vucetic, A. Dunker, and Z. Obradovic. 2006. Length-Dependent Prediction of Protein Intrinsic Disorder. *BMC bioinformatics*. 7:208, doi: 10.1186/1471-2105-7-208.

72. Dayhoff II, G. W., and V. N. Uversky. 2022. Rapid prediction and analysis of protein intrinsic disorder. *Protein Science*. 31(12):e4496, doi: https://doi.org/10.1002/pro.4496, https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4496.

73. Zhou, H.-X. 2004. Polymer Models of Protein Stability, Folding, and Interactions. *Biochemistry*. 43(8):2141-2154, doi: 10.1021/bi036269n, https://doi.org/10.1021/bi036269n.

74. Dill, K. A., S. Bromberg, K. Yue, H. S. Chan, K. M. Ftebig, D. P. Yee, and P. D. Thomas. 1995. Principles of protein folding — A perspective from simple exact models. *Protein Science*. 4(4):561-602, doi: https://doi.org/10.1002/pro.5560040401, https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5560040401.

75. Dill, K. A., and D. Stigter. 1995. Modeling Protein Stability As Heteropolymer Collapse. In Advances in Protein Chemistry. C. B. Anfinsen, F. M. Richards, J. T. Edsall, and D. S. Eisenberg, editors. Academic Press, pp. 59-104.

76. Dewey, T. G. 1993. Protein structure and polymer collapse. *The Journal of Chemical Physics*. 98(3):2250-2257, doi: 10.1063/1.464205, https://doi.org/10.1063/1.464205.

77. Dill, K. A. 1999. Polymer principles and protein folding. *Protein Science*. 8(6):1166-1180, doi: 10.1110/ps.8.6.1166, https://www.cambridge.org/core/article/polymer-principles-and-protein-folding/91C662F114689ED30432D6AF92E81D8C.

78. Milstein, J. N., and J.-C. Meiners. 2013. Worm-Like Chain (WLC) Model. In Encyclopedia of Biophysics. G. C. K. Roberts, editor. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 2757-2760.

79. O'Brien, E. P., G. Morrison, B. R. Brooks, and D. Thirumalai. 2009. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *The Journal of Chemical Physics*. 130(12), doi: 10.1063/1.3082151, https://doi.org/10.1063/1.3082151.

80. Banavar, J. R., T. X. Hoang, and A. Maritan. 2005. Proteins and polymers. *J Chem Phys*. 122(23):234910, doi: 10.1063/1.1940059, https://www.ncbi.nlm.nih.gov/pubmed/16008490.

81. Cheng, R. R., A. T. Hawk, and D. E. Makarov. 2013. Exploring the role of internal friction in the dynamics of unfolded proteins using simple polymer models. *The Journal of Chemical Physics*. 138(7), doi: 10.1063/1.4792206, https://doi.org/10.1063/1.4792206.

82. Vitalis, A., X. Wang, and R. V. Pappu. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys J*. 93(6):1923-1937, doi: 10.1529/biophysj.107.110080.

83. Kolinski, A. 2011. Lattice Polymers and Protein Models. In Multiscale Approaches to Protein Modeling: Structure Prediction, Dynamics, Thermodynamics and Macromolecular Assemblies. A. Kolinski, editor. Springer New York, New York, NY, pp. 1-20.

84. Phillips, J. L., M. E. Colvin, and S. Newsam. 2011. Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinformatics*. 12(1):445, doi: 10.1186/1471-2105-12-445, https://doi.org/10.1186/1471-2105-12-445.

85. Ghosh, K., J. Huihui, M. Phillips, and A. Haider. 2022. Rules of Physical Mathematics Govern Intrinsically Disordered Proteins. *Annual Review of Biophysics*. 51(1):355-376, doi: 10.1146/annurev-biophys-120221-095357, https://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-120221-095357.

86. Rodríguez-Cabello, J. C., J. Reguera, A. Girotti, F. J. Arias, and M. Alonso. 2006. Genetic Engineering of Protein-Based Polymers: The Example of Elastinlike Polymers. In Ordered Polymeric Nanostructures at Surfaces. G. J. Vancso, editor. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119-167.

87. Frandsen, J. L., and H. Ghandehari. 2012. Recombinant protein-based polymers for advanced drug delivery. *Chemical Society Reviews*. 41(7):2696-2706, doi: 10.1039/C2CS15303C, http://dx.doi.org/10.1039/C2CS15303C (10.1039/C2CS15303C).

88. Numata, K. 2020. How to define and study structural proteins as biopolymer materials. *Polymer Journal*. 52(9):1043-1056, doi: 10.1038/s41428-020-0362-5, https://doi.org/10.1038/s41428-020-0362-5.

89. Prusiner, S. B. 1998. Prions. *Proc Natl Acad Sci U S A*. 95(23):13363-13383, doi: 10.1073/pnas.95.23.13363.

90. Prusiner, S. B. 1982. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science*. 216(4542):136-144, doi: doi:10.1126/science.6801762, https://www.science.org/doi/abs/10.1126/science.6801762.

91. Prusiner, S. B. 2001. Neurodegenerative Diseases and Prions. *New England Journal of Medicine*. 344(20):1516-1526, doi: 10.1056/nejm200105173442006, https://www.nejm.org/doi/full/10.1056/NEJM200105173442006.

92. Aguzzi, A., and A. M. Calella. 2009. Prions: Protein Aggregation and Infectious Diseases. *Physiological Reviews*. 89(4):1105-1152, doi: 10.1152/physrev.00006.2009, https://journals.physiology.org/doi/abs/10.1152/physrev.00006.2009.

93. Aguzzi, A., M. Heikenwalder, and M. Polymenidou. 2007. Insights into prion strains and neurotoxicity. *Nature Reviews Molecular Cell Biology*. 8(7):552-561, doi: 10.1038/nrm2204, https://doi.org/10.1038/nrm2204.

94. Srivastava, K. R., and L. J. Lapidus. 2017. Prion protein dynamics before aggregation. *Proceedings of the National Academy of Sciences*. 114(14):3572-3577, doi: doi:10.1073/pnas.1620400114, https://www.pnas.org/doi/abs/10.1073/pnas.1620400114.

95. Goedert, M., F. Clavaguera, and M. Tolnay. 2010. The propagation of prion-like protein inclusions in neurodegenerative diseases. *Trends in Neurosciences*. 33(7):317-325, doi: https://doi.org/10.1016/j.tins.2010.04.003, https://www.sciencedirect.com/science/article/pii/S016622361000055X.

96. Ross, C. A., and M. A. Poirier. 2004. Protein aggregation and neurodegenerative disease. *Nature Medicine*. 10(7):S10-S17, doi: 10.1038/nm1066, https://doi.org/10.1038/nm1066.

97. Yi, C.-W., W.-C. Xu, J. Chen, and Y. Liang. 2013. Recent progress in prion and prion-like protein aggregation. *Acta Biochimica et Biophysica Sinica*. 45(6):520-526, doi: 10.1093/abbs/gmt052, https://doi.org/10.1093/abbs/gmt052.

98. Gil-Garcia, M., V. Iglesias, I. Pallarès, and S. Ventura. 2021. Prion-like proteins: from computational approaches to proteome-wide analysis. *FEBS Open Bio*. 11(9):2400-2417, doi: 10.1002/2211-5463.13213.

99. Yushchenko, T., E. Deuerling, and K. Hauser. 2018. Insights into the Aggregation Mechanism of PolyQ Proteins with Different Glutamine Repeat Lengths. *Biophys J*. 114(8):1847-1857, doi: 10.1016/j.bpj.2018.02.037.

100. Siu, H.-W., P. Stritt, H. Zhao, and K. Hauser. 2022. PolyQ aggregation studied by model peptides with intrinsic tryptophan fluorophores. *Biophysical Chemistry*. 284:106782, doi: 10.1016/j.bpc.2022.106782, https://dx.doi.org/10.1016/j.bpc.2022.106782.

101. Sodek, J., R. S. Hodges, L. B. Smillie, and L. Jurasek. 1972. Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure. *Proc Natl Acad Sci U S A*. 69(12):3800-3804, doi: 10.1073/pnas.69.12.3800.

102. Fiumara, F., L. Fioriti, E. R. Kandel, and W. A. Hendrickson. 2010. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell*. 143(7):1121-1135, doi: 10.1016/j.cell.2010.11.042.

103. Derkatch, I. L., S. M. Uptain, T. F. Outeiro, R. Krishnan, S. L. Lindquist, and S. W. Liebman. 2004. Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI⁺] prion in yeast and aggregation of Sup35 in vitro. *Proceedings of the National Academy of Sciences*. 101(35):12934-12939, doi: doi:10.1073/pnas.0404968101, https://www.pnas.org/doi/abs/10.1073/pnas.0404968101.

104. Chan, J. C., N. A. Oyler, W. M. Yau, and R. Tycko. 2005. Parallel beta-sheets and polar zippers in amyloid fibrils formed by residues 10-39 of the yeast prion protein Ure2p. *Biochemistry*. 44(31):10669-10680, doi: 10.1021/bi050724t.

105. Ross, E. D., A. Minton, and R. B. Wickner. 2005. Prion domains: sequences, structures and interactions. *Nat Cell Biol*. 7(11):1039-1044, doi: 10.1038/ncb1105-1039.

106. Gazi, A. D., M. Bastaki, S. N. Charova, E. A. Gkougkoulia, E. A. Kapellios, N. J. Panopoulos, and M. Kokkinidis. 2008. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. *J Biol Chem*. 283(49):34062-34068, doi: 10.1074/jbc.M803408200, https://www.ncbi.nlm.nih.gov/pubmed/18836182.

107. Seim, I., A. E. Posey, W. T. Snead, B. M. Stormo, D. Klotsa, R. V. Pappu, and A. S. Gladfelter (2021). Negative regulation of a ribonucleoprotein condensate driven by dilute phase oligomerization. Cold Spring Harbor Laboratory.

108. Zhang, Y., S. Li, X. Gong, and J. Chen. 2024. Toward Accurate Simulation of Coupling between Protein Secondary Structure and Phase Separation. *Journal of the American Chemical Society*. 146(1):342-357, doi: 10.1021/jacs.3c09195, https://doi.org/10.1021/jacs.3c09195.

109. Wickner, R. B., H. K. Edskes, F. Shewmaker, and T. Nakayashiki. 2007. Prions of fungi: inherited structures and biological roles. *Nat Rev Microbiol*. 5(8):611-618, doi: 10.1038/nrmicro1708.

110. Zbinden, A., M. Pérez-Berlanga, P. De Rossi, and M. Polymenidou. 2020. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Dev Cell*. 55(1):45-68, doi: 10.1016/j.devcel.2020.09.014.

111. Franzmann, T. M., M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman, and S. Alberti. 2018. Phase separation of a yeast prion protein promotes cellular fitness. *Science*. 359(6371), doi: 10.1126/science.aao5654.

112. March, Z. M., O. D. King, and J. Shorter. 2016. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Research*. 1647:9-18, doi: https://doi.org/10.1016/j.brainres.2016.02.037, https://www.sciencedirect.com/science/article/pii/S0006899316300968.

113. Martin, E. W., A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science*. 367(6478):694-699, doi: 10.1126/science.aaw8653, https://dx.doi.org/10.1126/science.aaw8653.

114. Gotor, N. L., A. Armaos, G. Calloni, M. Torrent Burgas, R. M. Vabulas, N. S. De Groot, and G. G. Tartaglia. 2020. RNA-binding and prion domains: the Yin and Yang of phase separation. *Nucleic Acids Res*. 48(17):9491-9504, doi: 10.1093/nar/gkaa681.

115. Zhao, D., F. J. B. Bäuerlein, I. Saha, U. Hartl, W. Baumeister, and F. Wilfling (2023). Autophagy preferentially degrades non-fibrillar polyQ aggregates. bioRxiv.

116. Hutin, S., J. R. Kumita, V. I. Strotmann, A. Dolata, W. L. Ling, N. Louafi, A. Popov, P.-E. Milhiet, M. Blackledge, M. H. Nanao, P. A. Wigge, Y. Stahl, L. Costa, M. D. Tully, and C. Zubieta. 2023. Phase separation and molecular ordering of the prion-like domain of the Arabidopsis thermosensory protein EARLY FLOWERING 3. *Proceedings of the National Academy of Sciences*. 120(28):e2304714120, doi: doi:10.1073/pnas.2304714120, https://www.pnas.org/doi/abs/10.1073/pnas.2304714120.

117. Gutierrez, J. I., G. P. Brittingham, Y. Karadeniz, K. D. Tran, A. Dutta, A. S. Holehouse, C. L. Peterson, and L. J. Holt. 2022. SWI/SNF senses carbon starvation with a pH-sensitive low-complexity sequence. *eLife*. 11, doi: 10.7554/elife.70344, https://dx.doi.org/10.7554/elife.70344.

118. Verma, M., A. Girdhar, B. Patel, N. K. Ganguly, R. Kukreti, and V. Taneja. 2018. Q-Rich Yeast Prion [PSI+] Accelerates Aggregation of Transthyretin, a Non-Q-

Rich Human Protein. *Frontiers in Molecular Neuroscience*. 11, doi: 10.3389/fnmol.2018.00075, https://www.frontiersin.org/articles/10.3389/fnmol.2018.00075 (Original Research).

119. Chen, Z., X. Xiao, L. Yang, C. Lian, S. Xu, and H. Liu. 2023. Prion-like Aggregation of the Heptapeptide GNNQQNY into Amyloid Nanofiber Is Governed by Configuration Entropy. *Journal of Chemical Information and Modeling*. doi: 10.1021/acs.jcim.3c00370, https://doi.org/10.1021/acs.jcim.3c00370.

120. Duennwald, M. L., S. Jagadish, F. Giorgini, P. J. Muchowski, and S. Lindquist. 2006. A network of protein interactions determines polyglutamine toxicity. *Proceedings of the National Academy of Sciences*. 103(29):11051-11056, doi: doi:10.1073/pnas.0604548103, https://www.pnas.org/doi/abs/10.1073/pnas.0604548103.

121. Gruber, A., D. Hornburg, M. Antonin, N. Krahmer, J. Collado, M. Schaffer, G. Zubaite, C. Lüchtenborg, T. Sachsenheimer, B. Brügger, M. Mann, W. Baumeister, F. U. Hartl, M. S. Hipp, and R. Fernández-Busnadiego. 2018. Molecular and structural architecture of polyQ aggregates in yeast. *Proceedings of the National Academy of Sciences*. 115(15):E3446-E3453, doi: doi:10.1073/pnas.1717978115, https://www.pnas.org/doi/abs/10.1073/pnas.1717978115.

122. Ripaud, L., V. Chumakova, M. Antonin, A. R. Hastie, S. Pinkert, R. Körner, K. M. Ruff, R. V. Pappu, D. Hornburg, M. Mann, F. U. Hartl, and M. S. Hipp. 2014. Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. *Proceedings of the National Academy of Sciences*. 111(51):18219-18224, doi: 10.1073/pnas.1421313111, https://dx.doi.org/10.1073/pnas.1421313111.

123. Bolognesi, B., A. J. Faure, M. Seuma, J. M. Schmiedel, G. G. Tartaglia, and B. Lehner. 2019. The mutational landscape of a prion-like domain. *Nature Communications*. 10(1):4162, doi: 10.1038/s41467-019-12101-z, https://doi.org/10.1038/s41467-019-12101-z.

124. Kaganovich, D., R. Kopito, and J. Frydman. 2008. Misfolded proteins partition between two distinct quality control compartments. *Nature*. 454(7208):1088-1095, doi: 10.1038/nature07195, https://doi.org/10.1038/nature07195.

125. Kayatekin, C., K. E. S. Matlack, W. R. Hesse, Y. Guan, S. Chakrabortee, J. Russ, E. E. Wanker, J. V. Shah, and S. Lindquist. 2014. Prion-like proteins sequester and suppress the toxicity of huntingtin exon 1. *Proceedings of the National Academy of Sciences*. 111(33):12085-12090, doi: doi:10.1073/pnas.1412504111, https://www.pnas.org/doi/abs/10.1073/pnas.1412504111.

126. Mateju, D., T. M. Franzmann, A. Patel, A. Kopach, E. E. Boczek, S. Maharana, H. O. Lee, S. Carra, A. A. Hyman, and S. Alberti. 2017. An aberrant phase transition of stress granules triggered by misfolded protein and prevented by chaperone function. *Embo j*. 36(12):1669-1687, doi: 10.15252/embj.201695957.

127. Sanz-Hernández, M., J. D. Barritt, J. Sobek, S. Hornemann, A. Aguzzi, and A. De Simone. 2021. Mechanism of misfolding of the human prion protein revealed by a pathological mutation. *Proceedings of the National Academy of Sciences*. 118(12):e2019631118, doi: doi:10.1073/pnas.2019631118, https://www.pnas.org/doi/abs/10.1073/pnas.2019631118.

128. Mathieu, C., R. V. Pappu, and J. P. Taylor. 2020. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science*. 370(6512):56-60, doi: 10.1126/science.abb8032.

129. Patel, A., H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, and S. Alberti. 2015. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. 162(5):1066-1077, doi: 10.1016/j.cell.2015.07.047.

130. Meisl, G., T. Kurt, I. Condado-Morales, C. Bett, S. Sorce, M. Nuvolone, T. C. T. Michaels, D. Heinzer, M. Avar, S. I. A. Cohen, S. Hornemann, A. Aguzzi, C. M. Dobson, C. J. Sigurdson, and T. P. J. Knowles. 2021. Scaling analysis reveals the mechanism and rates of prion replication in vivo. *Nature Structural & Molecular Biology*. 28(4):365-372, doi: 10.1038/s41594-021-00565-x, https://doi.org/10.1038/s41594-021-00565-x.

131. Legname, G., and C. Scialò. 2020. On the role of the cellular prion protein in the uptake and signaling of pathological aggregates in neurodegenerative diseases. *Prion*. 14(1):257-270, doi: 10.1080/19336896.2020.1854034, https://doi.org/10.1080/19336896.2020.1854034.

132. Mudher, A., M. Colin, S. Dujardin, M. Medina, I. Dewachter, S. M. Alavi Naini, E. M. Mandelkow, E. Mandelkow, L. Buée, M. Goedert, and J. P. Brion. 2017. What is the evidence that tau pathology spreads through prion-like propagation? *Acta neuropathologica communications*. 5(1):99, doi: 10.1186/s40478-017-0488-7, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045708944&doi=10.1186%2fs40478-017-0488-7&partnerID=40&md5=e910c2d89f3a7bebd3c3861814c8017c (Review).

133. Costanzo, M., and C. Zurzolo. 2013. The cell biology of prion-like spread of protein aggregates: mechanisms and implication in neurodegeneration. *Biochemical Journal*. 452(1):1-17, doi: 10.1042/bj20121898, https://doi.org/10.1042/BJ20121898.

134. Polymenidou, M., and D. W. Cleveland. 2012. Prion-like spread of protein aggregates in neurodegeneration. *Journal of Experimental Medicine*. 209(5):889-893, doi: 10.1084/jem.20120741, https://doi.org/10.1084/jem.20120741.

135. Miyazawa, K., H. Okada, K. Masujin, Y. Iwamaru, and T. Yokoyama. 2015. Infectivity-associated PrPSc and disease duration-associated PrPSc of mouse BSE prions. *Prion*. 9(5):394-403, doi: 10.1080/19336896.2015.1111507, https://doi.org/10.1080/19336896.2015.1111507.

136. Corsaro, A., S. Thellung, V. Villa, M. Nizzari, and T. Florio. 2012. Role of Prion Protein Aggregation in Neurotoxicity. *International Journal of Molecular Sciences*. 13(7):8648-8669, https://www.mdpi.com/1422-0067/13/7/8648.

137. Block, A. J., and J. C. Bartz. 2023. Prion strains: shining new light on old concepts. *Cell and Tissue Research*. 392(1):113-133, doi: 10.1007/s00441-022-03665-2, https://doi.org/10.1007/s00441-022-03665-2.

138. Willbold, D., B. Strodel, G. F. Schröder, W. Hoyer, and H. Heise. 2021. Amyloid-type Protein Aggregation and Prion-like Properties of Amyloids. *Chemical Reviews*. 121(13):8285-8307, doi: 10.1021/acs.chemrev.1c00196, https://doi.org/10.1021/acs.chemrev.1c00196.

139. Meisl, G., T. P. J. Knowles, and D. Klenerman. 2020. The molecular processes underpinning prion-like spreading and seed amplification in protein aggregation. *Current Opinion in Neurobiology*. 61:58-64, doi: https://doi.org/10.1016/j.conb.2020.01.010, https://www.sciencedirect.com/science/article/pii/S0959438820300295.

140. Lambert, S. A., A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch. 2018. The Human Transcription Factors. *Cell*. 172(4):650-665, doi: 10.1016/j.cell.2018.01.029.

141. Rauluseviciute, I., R. Riudavets-Puig, R. Blanc-Mathieu, Jaime A. Castro-Mondragon, K. Ferenc, V. Kumar, R. B. Lemma, J. Lucas, J. Chèneby, D. Baranasic, A. Khan, O. Fornes, S. Gundersen, M. Johansen, E. Hovig, B. Lenhard, A. Sandelin, Wyeth W. Wasserman, F. Parcy, and A. Mathelier. 2023. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. 52(D1):D174-D182, doi: 10.1093/nar/gkad1059, https://doi.org/10.1093/nar/gkad1059.

142. Fornes, O., J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranašić, W. Santana-Garcia, G. Tan, J. Chèneby, B. Ballester, F. Parcy, A. Sandelin, B. Lenhard, W. W. Wasserman, and A. Mathelier. 2019. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. 48(D1):D87-D92, doi: 10.1093/nar/gkz1001, https://doi.org/10.1093/nar/gkz1001.

143. Koo, P. K., and M. Ploenzke. 2020. Deep learning for inferring transcription factor binding sites. *Current Opinion in Systems Biology*. 19:16-23, doi: https://doi.org/10.1016/j.coisb.2020.04.001, https://www.sciencedirect.com/science/article/pii/S2452310020300032.

144. Latchman, D. S. 1997. Transcription factors: An overview. *The International Journal of Biochemistry & Cell Biology*. 29(12):1305-1312, doi: https://doi.org/10.1016/S1357-2725(97)00085-X, https://www.sciencedirect.com/science/article/pii/S135727259700085X.

145. Papavassiliou, A. G. 1995. Transcription Factors. *New England Journal of Medicine*. 332(1):45-47, doi: 10.1056/nejm199501053320108, https://www.nejm.org/doi/full/10.1056/NEJM199501053320108.

146. Dai, C., X. Chen, J. Li, P. Comish, R. Kang, and D. Tang. 2020. Transcription factors in ferroptotic cell death. *Cancer Gene Therapy*. 27(9):645-656, doi: 10.1038/s41417-020-0170-2, https://doi.org/10.1038/s41417-020-0170-2.

147. Kaltschmidt, B., P. A. Baeuerle, and C. Kaltschmidt. 1993. Potential involvement of the transcription factor NF-κB in neurological disorders. *Molecular Aspects of Medicine*. 14(3):171-190, doi: https://doi.org/10.1016/0098-2997(93)90004-W, https://www.sciencedirect.com/science/article/pii/009829979390004W.

148. Santos-Terra, J., I. Deckmann, M. Fontes-Dutra, G. B. Schwingel, V. Bambini-Junior, and C. Gottfried. 2021. Transcription factors in neurodevelopmental and associated psychiatric disorders: A potential convergence for genetic and environmental risk factors. *International Journal of Developmental Neuroscience*. 81(7):545-578, doi: https://doi.org/10.1002/jdn.10141, https://onlinelibrary.wiley.com/doi/abs/10.1002/jdn.10141.

149. Estruch, S. B., S. A. Graham, M. Quevedo, A. Vino, D. H. W. Dekkers, P. Deriziotis, E. Sollis, J. Demmers, R. A. Poot, and S. E. Fisher. 2018. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. *Human Molecular Genetics*. 27(7):1212-1227, doi: 10.1093/hmg/ddy035, https://doi.org/10.1093/hmg/ddy035.

150. Co, M., A. G. Anderson, and G. Konopka. 2020. FOXP transcription factors in vertebrate brain development, function, and disorders. *WIREs Developmental Biology*. 9(5):e375, doi: https://doi.org/10.1002/wdev.375, https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wdev.375.

151. Zaret, K. S. 2020. Pioneer Transcription Factors Initiating Gene Network Changes. *Annual Review of Genetics*. 54(1):367-385, doi: 10.1146/annurev-genet-030220-015007, https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-030220-015007.

152. Vierstra, J., J. Lazar, R. Sandstrom, J. Halow, K. Lee, D. Bates, M. Diegel, D. Dunn, F. Neri, E. Haugen, E. Rynes, A. Reynolds, J. Nelson, A. Johnson, M. Frerker, M. Buckley, R. Kaul, W. Meuleman, and J. A. Stamatoyannopoulos. 2020. Global reference mapping of human transcription factor footprints. *Nature*. 583(7818):729-736, doi: 10.1038/s41586-020-2528-x, https://doi.org/10.1038/s41586-020-2528-x.

153. Oksuz, O., J. E. Henninger, R. Warneford-Thomson, M. M. Zheng, H. Erb, A. Vancura, K. J. Overholt, S. W. Hawken, S. F. Banani, R. Lauman, L. N. Reich, A. L. Robertson, N. M. Hannett, T. I. Lee, L. I. Zon, R. Bonasio, and R. A. Young. 2023. Transcription factors interact with RNA to regulate genes. *Molecular Cell*. 83(14):2449-2463.e2413, doi: https://doi.org/10.1016/j.molcel.2023.06.012, https://www.sciencedirect.com/science/article/pii/S1097276523004343.

154. Bartosovic, M., M. Kabbe, and G. Castelo-Branco. 2021. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology*. 39(7):825-835, doi: 10.1038/s41587-021-00869-9, https://doi.org/10.1038/s41587-021-00869-9.

155. Song, J.-X., J. Liu, Y. Jiang, Z.-Y. Wang, and M. Li. 2021. Transcription factor EB: an emerging drug target for neurodegenerative disorders. *Drug Discovery Today*. 26(1):164-172, doi: https://doi.org/10.1016/j.drudis.2020.10.013, https://www.sciencedirect.com/science/article/pii/S135964462030430X.

156. Teixeira, J. R., R. A. Szeto, V. M. A. Carvalho, A. R. Muotri, and F. Papes. 2021. Transcription factor 4 and its association with psychiatric disorders. *Translational Psychiatry*. 11(1):19, doi: 10.1038/s41398-020-01138-0, https://doi.org/10.1038/s41398-020-01138-0.

157. Zhou, X., W. Tang, T. A. Greenwood, S. Guo, L. He, M. A. Geyer, and J. R. Kelsoe. 2009. Transcription Factor SP4 Is a Susceptibility Gene for Bipolar Disorder. *PLOS ONE*. 4(4):e5196, doi: 10.1371/journal.pone.0005196, https://doi.org/10.1371/journal.pone.0005196.

158. Gharani, N., R. Benayed, V. Mancuso, L. M. Brzustowicz, and J. H. Millonig. 2004. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Molecular Psychiatry*. 9(5):474-484, doi: 10.1038/sj.mp.4001498, https://doi.org/10.1038/sj.mp.4001498.

159. De Francesco, E., M. Terzaghi, E. Storelli, L. Magistrelli, C. Comi, M. Legnaro, M. Mauri, F. Marino, M. Versino, and M. Cosentino. 2021. CD4+ T-cell Transcription Factors in Idiopathic REM Sleep Behavior Disorder and Parkinson's Disease. *Movement Disorders*. 36(1):225-229, doi: https://doi.org/10.1002/mds.28137, https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.28137.

160. Jiramongkol, Y., and E. W. F. Lam. 2020. FOXO transcription factor family in cancer and metastasis. *Cancer and Metastasis Reviews*. 39(3):681-709, doi: 10.1007/s10555-020-09883-w, https://doi.org/10.1007/s10555-020-09883-w.

161. CALKHOVEN, C. F., and G. AB. 1996. Multiple steps in the regulation of transcription-factor level and activity. *Biochemical Journal*. 317(2):329-342, doi: 10.1042/bj3170329, https://doi.org/10.1042/bj3170329.

162. Liu, J., N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. 2006. Intrinsic Disorder in Transcription Factors. *Biochemistry*. 45(22):6873-6888, doi: 10.1021/bi0602718, https://doi.org/10.1021/bi0602718.

163. Yruela, I., C. J. Oldfield, K. J. Niklas, and A. K. Dunker. 2017. Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and

Organismic Complexity. *Genome Biology and Evolution*. 9(5):1248-1265, doi: 10.1093/gbe/evx073, https://doi.org/10.1093/gbe/evx073.

164. GUO, X., M. L. BULYK, and A. J. HARTEMINK. 2012. INTRINSIC DISORDER WITHIN AND FLANKING THE DNA-BINDING DOMAINS OF HUMAN TRANSCRIPTION FACTORS. In Biocomputing. pp. 104-115.

165. Campbell, K. M., A. R. Terrell, P. J. Laybourn, and K. J. Lumb. 2000. Intrinsic Structural Disorder of the C-Terminal Activation Domain from the bZIP Transcription Factor Fos. *Biochemistry*. 39(10):2708-2713, doi: 10.1021/bi9923555, https://doi.org/10.1021/bi9923555.

166. Hilser, V. J., and E. B. Thompson. 2011. Structural Dynamics, Intrinsic Disorder, and Allostery in Nuclear Receptors as Transcription Factors *. *Journal of Biological Chemistry*. 286(46):39675-39682, doi: 10.1074/jbc.R111.278929, https://doi.org/10.1074/jbc.R111.278929.

167. Brodsky, S., T. Jana, K. Mittelman, M. Chapal, D. K. Kumar, M. Carmi, and N. Barkai. 2020. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular Cell*. 79(3):459-471.e454, doi: https://doi.org/10.1016/j.molcel.2020.05.032, https://www.sciencedirect.com/science/article/pii/S109727652030352X.

168. Wagh, K., D. A. Stavreva, A. Upadhyaya, and G. L. Hager. 2023. Transcription Factor Dynamics: One Molecule at a Time. *Annual Review of Cell and Developmental Biology*. 39(1):277-305, doi: 10.1146/annurev-cellbio-022823-013847, https://www.annualreviews.org/doi/abs/10.1146/annurev-cellbio-022823-013847.

169. Marceau, A. H., C. M. Brison, S. Nerli, H. E. Arsenault, A. C. McShan, E. Chen, H.-W. Lee, J. A. Benanti, N. G. Sgourakis, and S. M. Rubin. 2019. An order-to-disorder structural switch activates the FoxM1 transcription factor. *eLife*. 8:e46131, doi: 10.7554/eLife.46131, https://doi.org/10.7554/eLife.46131.

170. Hannon, C. E., and M. B. Eisen. 2024. Intrinsic protein disorder is insufficient to drive subnuclear clustering in embryonic transcription factors. *eLife*. 12:RP88221, doi: 10.7554/eLife.88221, https://doi.org/10.7554/eLife.88221.

171. Brodsky, S., T. Jana, and N. Barkai. 2021. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Current Opinion in Structural Biology*. 71:110-115, doi:

https://doi.org/10.1016/j.sbi.2021.06.011, https://www.sciencedirect.com/science/article/pii/S0959440X21000956.

172. Schneider, N., F.-G. Wieland, D. Kong, A. A. M. Fischer, M. Hörner, J. Timmer, H. Ye, and W. Weber. 2021. Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Science Advances*. 7(1):eabd3568, doi: doi:10.1126/sciadv.abd3568, https://www.science.org/doi/abs/10.1126/sciadv.abd3568.

173. Kolimi, N., A. Pabbathi, N. Saikia, F. Ding, H. Sanabria, and J. Alper. 2021. Out-of-Equilibrium Biophysical Chemistry: The Case for Multidimensional, Integrated Single-Molecule Approaches. *The Journal of Physical Chemistry B*. 125(37):10404-10418, doi: 10.1021/acs.jpcb.1c02424, https://doi.org/10.1021/acs.jpcb.1c02424.

174. Tian, Y., J. Booth, E. Meehan, D. S. Jones, S. Li, and G. P. Andrews. 2013. Construction of Drug–Polymer Thermodynamic Phase Diagrams Using Flory–Huggins Interaction Theory: Identifying the Relevance of Temperature and Drug Weight Fraction to Phase Separation within Solid Dispersions. *Molecular Pharmaceutics*. 10(1):236-248, doi: 10.1021/mp300386v, https://doi.org/10.1021/mp300386v.

175. Zhao, Y., P. Inbar, H. P. Chokshi, A. W. Malick, and D. S. Choi. 2011. Prediction of the thermal phase diagram of amorphous solid dispersions by Flory-Huggins theory. *J Pharm Sci*. 100(8):3196-3207, doi: 10.1002/jps.22541.

176. Pajula, K., M. Taskinen, V.-P. Lehto, J. Ketolainen, and O. Korhonen. 2010. Predicting the Formation and Stability of Amorphous Small Molecule Binary Mixtures from Computationally Determined Flory−Huggins Interaction Parameter and Phase Diagram. *Molecular Pharmaceutics*. 7(3):795-804, doi: 10.1021/mp900304p, https://dx.doi.org/10.1021/mp900304p.

177. Ghitman, J., R. Stan, G. Vlasceanu, E. Vasile, and H. Iovu. 2019. Predicting the drug loading efficiency into hybrid nanocarriers based on PLGA-vegetable oil using molecular dynamic simulation approach and Flory-Huggins theory. *Journal of Drug Delivery Science and Technology*. 53:101203, doi: 10.1016/j.jddst.2019.101203, https://dx.doi.org/10.1016/j.jddst.2019.101203.

178. Potter, C. B., M. T. Davis, A. B. Albadarin, and G. M. Walker. 2018. Investigation of the Dependence of the Flory–Huggins Interaction Parameter on Temperature and Composition in a Drug–Polymer System. *Molecular*

Pharmaceutics. 15(11):5327-5335, doi: 10.1021/acs.molpharmaceut.8b00797, https://dx.doi.org/10.1021/acs.molpharmaceut.8b00797.

179. Kozuch, D., W. Zhang, and S. Milner. 2016. Predicting the Flory-Huggins χ Parameter for Polymers with Stiffness Mismatch from Molecular Dynamics Simulations. *Polymers*. 8(6):241, doi: 10.3390/polym8060241, https://dx.doi.org/10.3390/polym8060241.

180. Callaway, C. P., K. Hendrickson, N. Bond, S. M. Lee, P. Sood, and S. S. Jang. 2018. Molecular Modeling Approach to Determine the Flory-Huggins Interaction Parameter for Polymer Miscibility Analysis. *ChemPhysChem*. 19(13):1655-1664, doi: 10.1002/cphc.201701337, https://dx.doi.org/10.1002/cphc.201701337.

181. Beardsley, T. M., and M. W. Matsen. 2019. Calibration of the Flory-Huggins interaction parameter in field-theoretic simulations. *The Journal of Chemical Physics*. 150(17):174902, doi: 10.1063/1.5089217, https://dx.doi.org/10.1063/1.5089217.

182. Zhang, W., E. D. Gomez, and S. T. Milner. 2017. Predicting Flory-Huggins χ from Simulations. *Physical Review Letters*. 119(1), doi: 10.1103/physrevlett.119.017801, https://dx.doi.org/10.1103/physrevlett.119.017801.

183. Lin, Y.-H., J. Song, J. D. Forman-Kay, and H. S. Chan. 2017. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids*. 228:176-193, doi: 10.1016/j.molliq.2016.09.090, https://dx.doi.org/10.1016/j.molliq.2016.09.090.

184. Dignon, G. L., W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal. 2018. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology*. 14(1):e1005941, doi: 10.1371/journal.pcbi.1005941, https://dx.doi.org/10.1371/journal.pcbi.1005941.

185. Riniker, S., J. R. Allison, and W. F. van Gunsteren. 2012. On developing coarse-grained models for biomolecular simulation: a review. *Physical Chemistry Chemical Physics*. 14(36):12423-12430, doi: 10.1039/C2CP40934H, http://dx.doi.org/10.1039/C2CP40934H (10.1039/C2CP40934H).

186. Singh, N., and W. Li. 2019. Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications. *International Journal of Molecular Sciences*. 20(15):3774, https://www.mdpi.com/1422-0067/20/15/3774.

187. Tozzini, V. 2005. Coarse-grained models for proteins. *Current Opinion in Structural Biology*. 15(2):144-150, doi: https://doi.org/10.1016/j.sbi.2005.02.005, https://www.sciencedirect.com/science/article/pii/S0959440X05000515.

188. Wu, C., and J.-E. Shea. 2011. Coarse-grained models for protein aggregation. *Current Opinion in Structural Biology*. 21(2):209-220, doi: https://doi.org/10.1016/j.sbi.2011.02.002, https://www.sciencedirect.com/science/article/pii/S0959440X11000339.

189. Potoyan, D. A., A. Savelyev, and G. A. Papoian. 2013. Recent successes in coarse-grained modeling of DNA. *WIREs Computational Molecular Science*. 3(1):69-83, doi: https://doi.org/10.1002/wcms.1114, https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1114.

190. Knotts, T. A., IV, N. Rathore, D. C. Schwartz, and J. J. de Pablo. 2007. A coarse grain model for DNA. *The Journal of Chemical Physics*. 126(8), doi: 10.1063/1.2431804, https://doi.org/10.1063/1.2431804.

191. Kim, Y. C., and G. Hummer. 2008. Coarse-grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. *Journal of Molecular Biology*. 375(5):1416-1433, doi: https://doi.org/10.1016/j.jmb.2007.11.063, https://www.sciencedirect.com/science/article/pii/S0022283607015628.

192. Baschnagel, J., K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, and V. Tries. 2000. Bridging the Gap Between Atomistic and Coarse-Grained Models of Polymers: Status and Perspectives. In Viscoelasticity, Atomistic Models, Statistical Chemistry. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 41-156.

193. Khot, A., S. B. Shiring, and B. M. Savoie. 2019. Evidence of information limitations in coarse-grained models. *The Journal of Chemical Physics*. 151(24), doi: 10.1063/1.5129398, https://doi.org/10.1063/1.5129398.

194. Noid, W. G., J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. 2008. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics*. 128(24), doi: 10.1063/1.2938860, https://doi.org/10.1063/1.2938860.

195. Mioduszewski, Ł., and M. Cieplak. 2018. Disordered peptide chains in an α-C-based coarse-grained model. *Physical Chemistry Chemical Physics*. 20(28):19057-19070, doi: 10.1039/c8cp03309a, https://dx.doi.org/10.1039/C8CP03309A.

196. Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. 2007. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*. 111(27):7812-7824, doi: 10.1021/jp071097f, https://doi.org/10.1021/jp071097f.

197. Izvekov, S., P. W. Chung, and B. M. Rice. 2010. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *The Journal of Chemical Physics*. 133(6), doi: 10.1063/1.3464776, https://doi.org/10.1063/1.3464776.

198. Schilling, T. 2022. Coarse-grained modelling out of equilibrium. *Physics Reports*. 972:1-45, doi: https://doi.org/10.1016/j.physrep.2022.04.006, https://www.sciencedirect.com/science/article/pii/S0370157322001314.

199. Clementi, C. 2008. Coarse-grained models of protein folding: toy models or predictive tools? *Current Opinion in Structural Biology*. 18(1):10-15, doi: https://doi.org/10.1016/j.sbi.2007.10.005, https://www.sciencedirect.com/science/article/pii/S0959440X07001753.

200. Henrich, O., Y. A. Gutiérrez Fosado, T. Curk, and T. E. Ouldridge. 2018. Coarse-grained simulation of DNA using LAMMPS. *The European Physical Journal E*. 41(5):57, doi: 10.1140/epje/i2018-11669-8, https://doi.org/10.1140/epje/i2018-11669-8.

201. Thompson, A. P., H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. 2022. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*. 271:108171, doi: https://doi.org/10.1016/j.cpc.2021.108171, https://www.sciencedirect.com/science/article/pii/S0010465521002836.

202. Luo, C., and J.-U. Sommer. 2009. Coding coarse grained polymer model for LAMMPS and its application to polymer crystallization. *Computer Physics Communications*. 180(8):1382-1391, doi:

https://doi.org/10.1016/j.cpc.2009.01.028,
https://www.sciencedirect.com/science/article/pii/S0010465509000423.

203. Kmiecik, S., D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski. 2016. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews*. 116(14):7898-7936, doi: 10.1021/acs.chemrev.6b00163, https://doi.org/10.1021/acs.chemrev.6b00163.

204. Noid, W. G. 2013. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*. 139(9), doi: 10.1063/1.4818908, https://doi.org/10.1063/1.4818908.

205. Plimpton, S. 1995. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*. 117(1):1-19, doi: 10.1006/jcph.1995.1039, https://dx.doi.org/10.1006/jcph.1995.1039.

206. Li, K., X. Lin, and J. Greenberg. 2016. Software citation, reuse and metadata considerations: An exploratory study examining LAMMPS. *Proceedings of the Association for Information Science and Technology*. 53(1):1-10, doi: https://doi.org/10.1002/pra2.2016.14505301072, https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pra2.2016.14505301072.

207. Fischer, J., V. Natoli, and D. Richie (2006). Optimization of LAMMPS. 2006 HPCMP Users Group Conference (HPCMP-UGC'06).

208. Svaneborg, C. 2012. LAMMPS framework for dynamic bonding and an application modeling DNA. *Computer Physics Communications*. 183(8):1793-1802, doi: https://doi.org/10.1016/j.cpc.2012.03.005, https://www.sciencedirect.com/science/article/pii/S0010465512001105.

209. Mackay, F. E., S. T. T. Ollila, and C. Denniston. 2013. Hydrodynamic forces implemented into LAMMPS through a lattice-Boltzmann fluid. *Computer Physics Communications*. 184(8):2021-2031, doi: https://doi.org/10.1016/j.cpc.2013.03.024, https://www.sciencedirect.com/science/article/pii/S001046551300132X.

210. Paula Leite, R., and M. de Koning. 2019. Nonequilibrium free-energy calculations of fluids using LAMMPS. *Computational Materials Science*. 159:316-326, doi: https://doi.org/10.1016/j.commatsci.2018.12.029, https://www.sciencedirect.com/science/article/pii/S0927025618308085.

211. Freitas, R., M. Asta, and M. de Koning. 2016. Nonequilibrium free-energy calculation of solids using LAMMPS. *Computational Materials Science*. 112:333-341, doi: https://doi.org/10.1016/j.commatsci.2015.10.050, https://www.sciencedirect.com/science/article/pii/S0927025615007089.

212. FrantzDale, B., S. J. Plimpton, and M. S. Shephard. 2010. Software components for parallel multiscale simulation: an example with LAMMPS. *Engineering with Computers*. 26(2):205-211, doi: 10.1007/s00366-009-0156-z, https://doi.org/10.1007/s00366-009-0156-z.

213. Chaimovich, M., and A. Chaimovich. 2021. Relative Resolution: A Computationally Efficient Implementation in LAMMPS. *Journal of Chemical Theory and Computation*. 17(2):1045-1059, doi: 10.1021/acs.jctc.0c01003, https://doi.org/10.1021/acs.jctc.0c01003.

214. Dickel, D., M. Nitol, and C. D. Barrett. 2021. LAMMPS implementation of rapid artificial neural network derived interatomic potentials. *Computational Materials Science*. 196:110481, doi: https://doi.org/10.1016/j.commatsci.2021.110481, https://www.sciencedirect.com/science/article/pii/S0927025621002068.

215. Sudarsan, R., C. J. Ribbens, and D. Farkas (2009). Dynamic Resizing of Parallel Scientific Simulations: A Case Study Using LAMMPS. Springer Berlin Heidelberg.

216. Wu, G., Y. J. Bae, M. Olesińska, D. Antón-García, I. Szabó, E. Rosta, M. R. Wasielewski, and O. A. Scherman. 2019. Controlling the structure and photophysics of fluorophore dimers using multiple cucurbit[8]uril clampings. *Chem Sci*. 11(3):812-825, doi: 10.1039/c9sc04587b.

217. Chen, K., R. Yan, L. Xiang, and K. Xu. 2021. Excitation spectral microscopy for highly multiplexed fluorescence imaging and quantitative biosensing. *Light Sci Appl*. 10(1):97, doi: 10.1038/s41377-021-00536-3.

218. Xu, C., and W. W. Webb. 2002. Multiphoton Excitation of Molecular Fluorophores and Nonlinear Laser Microscopy. In Topics in Fluorescence Spectroscopy: Volume 5: Nonlinear and Two-Photon-Induced Fluorescence. J. R. Lakowicz, editor. Springer US, Boston, MA, pp. 471-540.

219. Lakowicz, J. R., and I. Gryczynski. 2002. Multiphoton Excitation of Biochemical Fluorophores. In Topics in Fluorescence Spectroscopy: Volume 5: Nonlinear and

Two-Photon-Induced Fluorescence. J. R. Lakowicz, editor. Springer US, Boston, MA, pp. 87-144.

220. Xu, C., R. M. Williams, W. Zipfel, and W. W. Webb. 1996. Multiphoton excitation cross-sections of molecular fluorophores. *Bioimaging*. 4(3):198-207, doi: https://doi.org/10.1002/1361-6374(199609)4:3<198::AID-BIO10>3.0.CO;2-X, https://onlinelibrary.wiley.com/doi/abs/10.1002/1361-6374%28199609%294%3A3%3C198%3A%3AAID-BIO10%3E3.0.CO%3B2-X.

221. Chorvat Jr., D., and A. Chorvatova. 2009. Multi-wavelength fluorescence lifetime spectroscopy: a new approach to the study of endogenous fluorescence in living cells and tissues. *Laser Physics Letters*. 6(3):175-193, doi: https://doi.org/10.1002/lapl.200810132, https://onlinelibrary.wiley.com/doi/abs/10.1002/lapl.200810132.

222. Gautheron, A., M. Sdika, M. Hébert, and B. Montcel. 2024. An Explicit Estimated Baseline Model for Robust Estimation of Fluorophores Using Multiple-Wavelength Excitation Fluorescence Spectroscopy. *IEEE Transactions on Biomedical Engineering*. 71(1):295-306, doi: 10.1109/TBME.2023.3299689.

223. Crisalli, P., and E. T. Kool. 2011. Multi-Path Quenchers: Efficient Quenching of Common Fluorophores. *Bioconjugate Chemistry*. 22(11):2345-2354, doi: 10.1021/bc200424r, https://doi.org/10.1021/bc200424r.

224. Chalfie, M. 1995. GREEN FLUORESCENT PROTEIN. *Photochemistry and Photobiology*. 62(4):651-656, doi: https://doi.org/10.1111/j.1751-1097.1995.tb08712.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-1097.1995.tb08712.x.

225. Remington, S. J. 2011. Green fluorescent protein: a perspective. *Protein Sci*. 20(9):1509-1519, doi: 10.1002/pro.684.

226. Follenius-Wund, A., M. Bourotte, M. Schmitt, F. Iyice, H. Lami, J.-J. Bourguignon, J. Haiech, and C. Pigault. 2003. Fluorescent Derivatives of the GFP Chromophore Give a New Insight into the GFP Fluorescence Process. *Biophysical Journal*. 85(3):1839-1850, doi: 10.1016/S0006-3495(03)74612-8, https://doi.org/10.1016/S0006-3495(03)74612-8.

227. Mukherjee, S., P. Manna, S.-T. Hung, F. Vietmeyer, P. Friis, A. E. Palmer, and R. Jimenez. 2022. Directed Evolution of a Bright Variant of mCherry: Suppression of Nonradiative Decay by Fluorescence Lifetime Selections. *The Journal of*

Physical Chemistry B. 126(25):4659-4668, doi: 10.1021/acs.jpcb.2c01956, https://doi.org/10.1021/acs.jpcb.2c01956.

228. Hartmann, A., G. Krainer, and M. Schlierf. 2014. Different fluorophore labeling strategies and designs affect millisecond kinetics of DNA hairpins. *Molecules*. 19(9):13735-13754, doi: 10.3390/molecules190913735.

229. Wimmi, S., A. Balinovic, C. Brianceau, K. Pintor, J. Vielhauer, B. Turkowyd, C. Helbig, M. Fleck, K. Langenfeld, J. Kahnt, T. Glatter, U. Endesfelder, and A. Diepold. 2024. Cytosolic sorting platform complexes shuttle type III secretion system effectors to the injectisome in Yersinia enterocolitica. *Nature Microbiology*. 9(1):185-199, doi: 10.1038/s41564-023-01545-1, https://doi.org/10.1038/s41564-023-01545-1.

230. Vojnovic, I., J. Winkelmeier, and U. Endesfelder. 2019. Visualizing the inner life of microbes: practices of multi-color single-molecule localization microscopy in microbiology. *Biochemical Society Transactions*. 47(4):1041-1065, doi: 10.1042/bst20180399, https://doi.org/10.1042/BST20180399.

231. Vojnovic, I., and U. Endesfelder. 2020. Beginner's guide to producing super-resolved images on a widefield fluorescence microscope. *The Biochemist*. 42(4):52-56, doi: 10.1042/bio20200045, https://doi.org/10.1042/BIO20200045.

232. Turkowyd, B., S. Schreiber, J. Wörtz, E. S. Segal, M. Mevarech, I. G. Duggin, A. Marchfelder, and U. Endesfelder. 2020. Establishing Live-Cell Single-Molecule Localization Microscopy Imaging and Single-Particle Tracking in the Archaeon Haloferax volcanii. *Frontiers in Microbiology*. 11, doi: 10.3389/fmicb.2020.583010, https://www.frontiersin.org/articles/10.3389/fmicb.2020.583010 (Methods).

233. Vangindertael, J., R. Camacho, W. Sempels, H. Mizuno, P. Dedecker, and K. P. F. Janssen. 2018. An introduction to optical super-resolution microscopy for the adventurous biologist. *Methods and Applications in Fluorescence*. 6(2):022003, doi: 10.1088/2050-6120/aaae0c, https://dx.doi.org/10.1088/2050-6120/aaae0c.

234. Schermelleh, L., A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier, and G. P. C. Drummen. 2019. Super-resolution microscopy demystified. *Nature Cell Biology*. 21(1):72-84, doi: 10.1038/s41556-018-0251-8, https://doi.org/10.1038/s41556-018-0251-8.

235. Galbraith, C. G., and J. A. Galbraith. 2011. Super-resolution microscopy at a glance. *Journal of Cell Science*. 124(10):1607-1611, doi: 10.1242/jcs.080085, https://doi.org/10.1242/jcs.080085.

236. Endesfelder, U., and M. Heilemann. 2015. Direct Stochastic Optical Reconstruction Microscopy (dSTORM). In Advanced Fluorescence Microscopy: Methods and Protocols. P. J. Verveer, editor. Springer New York, New York, NY, pp. 263-276.

237. Turkowyd, B., D. Virant, and U. Endesfelder. 2016. From single molecules to life: microscopy at the nanoscale. *Analytical and Bioanalytical Chemistry*. 408(25):6885-6911, doi: 10.1007/s00216-016-9781-8, https://doi.org/10.1007/s00216-016-9781-8.

238. Lakowicz, J. R. 2006. Principles of Fluorescence Spectroscopy. Springer US.

239. Agam, G., C. Gebhardt, M. Popara, R. Mächtel, J. Folz, B. Ambrose, N. Chamachi, S. Y. Chung, T. D. Craggs, M. de Boer, D. Grohmann, T. Ha, A. Hartmann, J. Hendrix, V. Hirschfeld, C. G. Hübner, T. Hugel, D. Kammerer, H.-S. Kang, A. N. Kapanidis, G. Krainer, K. Kramm, E. A. Lemke, E. Lerner, E. Margeat, K. Martens, J. Michaelis, J. Mitra, G. G. Moya Muñoz, R. B. Quast, N. C. Robb, M. Sattler, M. Schlierf, J. Schneider, T. Schröder, A. Sefer, P. S. Tan, J. Thurn, P. Tinnefeld, J. van Noort, S. Weiss, N. Wendler, N. Zijlstra, A. Barth, C. A. M. Seidel, D. C. Lamb, and T. Cordes. 2023. Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins. *Nature Methods*. 20(4):523-535, doi: 10.1038/s41592-023-01807-0, https://doi.org/10.1038/s41592-023-01807-0.

240. Sisamakis, E., A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. M. Seidel. 2010. Chapter 18 - Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. In Methods in Enzymology. N. G. Walter, editor. Academic Press, pp. 455-514.

241. Nettels, D., A. Hoffmann, and B. Schuler. 2008. Unfolded Protein and Peptide Dynamics Investigated with Single-Molecule FRET and Correlation Spectroscopy from Picoseconds to Seconds. *The Journal of Physical Chemistry B*. 112(19):6137-6146, doi: 10.1021/jp076971j, https://doi.org/10.1021/jp076971j.

242. Magde, D., E. Elson, and W. W. Webb. 1972. Thermodynamic Fluctuations in a Reacting System---Measurement by Fluorescence Correlation Spectroscopy.

Physical Review Letters. 29(11):705-708, doi: 10.1103/PhysRevLett.29.705, https://link.aps.org/doi/10.1103/PhysRevLett.29.705.

243. Elson, E. L., and D. Magde. 1974. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers*. 13(1):1-27, doi: https://doi.org/10.1002/bip.1974.360130102, https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1974.360130102.

244. Magde, D., W. W. Webb, and E. L. Elson. 1978. Fluorescence correlation spectroscopy. III. Uniform translation and laminar flow. *Biopolymers*. 17(2):361-376, doi: https://doi.org/10.1002/bip.1978.360170208, https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1978.360170208.

245. Sanabria, H., D. Rodnin, K. Hemmen, T.-O. Peulen, S. Felekyan, M. R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke, and C. A. M. Seidel. 2020. Resolving dynamics and function of transient states in single enzyme molecules. *Nature Communications*. 11(1):1231, doi: 10.1038/s41467-020-14886-w, https://doi.org/10.1038/s41467-020-14886-w.

246. Roy, R., S. Hohng, and T. Ha. 2008. A practical guide to single-molecule FRET. *Nature Methods*. 5(6):507-516, doi: 10.1038/nmeth.1208, https://doi.org/10.1038/nmeth.1208.

247. Hellenkamp, B., S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kühnemuth, S. Rezaei Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M. E. Bowen, H. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Götz, G. Gouridis, E. Gratton, T. Ha, P. Hao, C. A. Hanke, A. Hartmann, J. Hendrix, L. L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Hua, C. G. Hübner, E. Kallis, A. N. Kapanidis, J.-Y. Kim, G. Krainer, D. C. Lamb, N. K. Lee, E. A. Lemke, B. Levesque, M. Levitus, J. J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Qiu, N. C. Robb, C. Röcker, H. Sanabria, M. Schlierf, T. Schröder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberk, A. M. Vera, K. R. Weninger, B. Wünsch, I. S. Yanez-Orozco, J. Michaelis, C. A. M. Seidel, T. D. Craggs, and T. Hugel. 2018. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nature Methods*. 15(9):669-676, doi: 10.1038/s41592-018-0085-0, https://doi.org/10.1038/s41592-018-0085-0.

248. Algar, W. R., N. Hildebrandt, S. S. Vogel, and I. L. Medintz. 2019. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. *Nature Methods*. 16(9):815-829, doi: 10.1038/s41592-019-0530-8, https://doi.org/10.1038/s41592-019-0530-8.

249. Shrestha, D., A. Jenei, P. Nagy, G. Vereb, and J. Szöllősi. 2015. Understanding FRET as a Research Tool for Cellular Studies. *International Journal of Molecular Sciences*. 16(4):6718-6756, https://www.mdpi.com/1422-0067/16/4/6718.

250. Leavesley, S. J., and T. C. Rich. 2016. Overcoming limitations of FRET measurements. *Cytometry A*. 89(4):325-327, doi: 10.1002/cyto.a.22851.

251. Watrob, H. M., C.-P. Pan, and M. D. Barkley. 2003. Two-Step FRET as a Structural Tool. *Journal of the American Chemical Society*. 125(24):7336-7343, doi: 10.1021/ja034564p, https://doi.org/10.1021/ja034564p.

252. Vandermeer, B. W. 2020. Kappaphobia is the elephant in the fret room. *Methods and Applications in Fluorescence*. 8(3):030401, doi: 10.1088/2050-6120/ab8f87, https://dx.doi.org/10.1088/2050-6120/ab8f87.

253. Clegg, R. M. 2006. The History of Fret. In Reviews in Fluorescence 2006. C. D. Geddes, and J. R. Lakowicz, editors. Springer US, Boston, MA, pp. 1-45.

254. Durham, R. J., D. R. Latham, H. Sanabria, and V. Jayaraman. 2020. Structural Dynamics of Glutamate Signaling Systems by smFRET. *Biophysical Journal*. 119(10):1929-1936, doi: https://doi.org/10.1016/j.bpj.2020.10.009, https://www.sciencedirect.com/science/article/pii/S0006349520308110.

255. Lau, C. G., and R. S. Zukin. 2007. NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nature Reviews Neuroscience*. 8(6):413-426, doi: 10.1038/nrn2153, https://doi.org/10.1038/nrn2153.

256. Warnet, X. L., H. Bakke Krog, O. G. Sevillano-Quispe, H. Poulsen, and M. Kjaergaard. 2021. The C-terminal domains of the NMDA receptor: How intrinsically disordered tails affect signalling, plasticity and disease. *European Journal of Neuroscience*. 54(8):6713-6739, doi: https://doi.org/10.1111/ejn.14842, https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.14842.

257. Kutsuwada, T., K. Sakimura, T. Manabe, C. Takayama, N. Katakura, E. Kushiya, R. Natsume, M. Watanabe, Y. Inoue, T. Yagi, S. Aizawa, M. Arakawa, T. Takahashi, Y. Nakamura, H. Mori, and M. Mishina. 1996. Impairment of suckling response, trigeminal neuronal pattern formation, and hippocampal LTD in NMDA receptor epsilon 2 subunit mutant mice. *Neuron*. 16(2):333-344, doi: 10.1016/s0896-6273(00)80051-3.

258. Punnakkal, P., P. Jendritza, and G. Köhr. 2012. Influence of the intracellular GluN2 C-terminal domain on NMDA receptor function. *Neuropharmacology*. 62(5):1985-1992, doi: https://doi.org/10.1016/j.neuropharm.2011.12.018, https://www.sciencedirect.com/science/article/pii/S0028390811005752.

259. Choi, U. B., S. Xiao, L. P. Wollmuth, and M. E. Bowen. 2011. Effect of Src Kinase Phosphorylation on Disordered C-terminal Domain of N-Methyl-D-aspartic Acid (NMDA) Receptor Subunit GluN2B Protein. 286(34):29904-29912, doi: 10.1074/jbc.m111.258897, https://dx.doi.org/10.1074/jbc.M111.258897.

260. Chatterjee, S., C. Ade, C. E. Nurik, N. C. Carrejo, C. Dutta, V. Jayaraman, and C. F. Landes. 2019. Phosphorylation Induces Conformational Rigidity at the C-Terminal Domain of AMPA Receptors. *The Journal of Physical Chemistry B*. 123(1):130-137, doi: 10.1021/acs.jpcb.8b10749, https://doi.org/10.1021/acs.jpcb.8b10749.

261. Choi, U. B., R. Kazi, N. Stenzoski, L. P. Wollmuth, V. N. Uversky, and M. E. Bowen. 2013. Modulating the Intrinsic Disorder in the Cytoplasmic Domain Alters the Biological Activity of the N-Methyl-d-aspartate-sensitive Glutamate Receptor *. *Journal of Biological Chemistry*. 288(31):22506-22515, doi: 10.1074/jbc.M113.477810, https://doi.org/10.1074/jbc.M113.477810.

262. Evers, T. H., E. M. W. M. van Dongen, A. C. Faesen, E. W. Meijer, and M. Merkx. 2006. Quantitative Understanding of the Energy Transfer between Fluorescent Proteins Connected via Flexible Peptide Linkers. *Biochemistry*. 45(44):13183-13192, doi: 10.1021/bi061288t, https://doi.org/10.1021/bi061288t.

263. Choi, Ucheor B., James J. McCann, Keith R. Weninger, and Mark E. Bowen. 2011. Beyond the Random Coil: Stochastic Conformational Switching in Intrinsically Disordered Proteins. *Structure*. 19(4):566-576, doi: https://doi.org/10.1016/j.str.2011.01.011, https://www.sciencedirect.com/science/article/pii/S096921261100061X.

264. Choi, U. B., H. Sanabria, T. Smirnova, M. E. Bowen, and K. R. Weninger. 2019. Spontaneous Switching among Conformational Ensembles in Intrinsically Disordered Proteins. *Biomolecules*. 9(3):114, https://www.mdpi.com/2218-273X/9/3/114.

265. Olofsson, L., S. Felekyan, E. Doumazane, P. Scholler, L. Fabre, J. M. Zwier, P. Rondard, C. A. M. Seidel, J. P. Pin, and E. Margeat. 2014. Fine tuning of sub-millisecond conformational dynamics controls metabotropic glutamate receptors agonist efficacy. *Nature Communications*. 5, 5206, doi: 10.1038/ncomms6206, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84948949784&doi=10.1038%2fncomms6206&partnerID=40&md5=681c3544725100fc2af4b6895e96ca65 (Article).

266. Kudryavtsev, V., M. Sikor, S. Kalinin, D. Mokranjac, C. A. M. Seidel, and D. C. Lamb. 2012. Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem*. 13(4):1060-1078, doi: https://doi.org/10.1002/cphc.201100822, https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cphc.201100822.

267. Hamilton, G. L., N. Saikia, S. Basak, F. S. Welcome, F. Wu, J. Kubiak, C. Zhang, Y. Hao, C. A. M. Seidel, F. Ding, H. Sanabria, and M. E. Bowen. 2022. Fuzzy supertertiary interactions within PSD-95 enable ligand binding. *eLife*. 11:e77242, doi: 10.7554/eLife.77242, https://doi.org/10.7554/eLife.77242.

268. Yanez Orozco, I. S., F. A. Mindlin, J. Ma, B. Wang, B. Levesque, M. Spencer, S. Rezaei Adariani, G. Hamilton, F. Ding, M. E. Bowen, and H. Sanabria. 2018. Identifying weak interdomain interactions that stabilize the supertertiary structure of the N-terminal tandem PDZ domains of PSD-95. *Nature Communications*. 9(1), doi: 10.1038/s41467-018-06133-0, https://dx.doi.org/10.1038/s41467-018-06133-0.

269. Maus, M., M. Cotlet, J. Hofkens, T. Gensch, F. C. De Schryver, J. Schaffer, and C. A. M. Seidel. 2001. An Experimental Comparison of the Maximum Likelihood Estimation and Nonlinear Least-Squares Fluorescence Lifetime Analysis of Single Molecules. *Analytical Chemistry*. 73(9):2078-2086, doi: 10.1021/ac000877g, https://doi.org/10.1021/ac000877g.

270. Jernigan, R. L., and P. J. Flory. 1969. Distribution Functions for Chain Molecules. *The Journal of Chemical Physics*. 50(10):4185-4200, doi: 10.1063/1.1670884, https://doi.org/10.1063/1.1670884.

271. Grosberg, A. Y., T. T. Nguyen, and B. I. Shklovskii. 2002. Colloquium: The physics of charge inversion in chemical and biological systems. *Reviews of Modern Physics*. 74(2):329-345, doi: 10.1103/RevModPhys.74.329, https://link.aps.org/doi/10.1103/RevModPhys.74.329.

272. Basak, S., N. Saikia, D. Kwun, U. B. Choi, F. Ding, and M. E. Bowen. 2023. Different Forms of Disorder in NMDA-Sensitive Glutamate Receptor Cytoplasmic Domains Are Associated with Differences in Condensate Formation. *Biomolecules*. 13(1):4, https://www.mdpi.com/2218-273X/13/1/4.

273. Aguzzi, A., and M. Altmeyer. 2016. Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends in Cell Biology*. 26(7):547-558, doi: 10.1016/j.tcb.2016.03.004, https://dx.doi.org/10.1016/j.tcb.2016.03.004.

274. Boersma, A. J., I. S. Zuhorn, and B. Poolman. 2015. A sensor for quantification of macromolecular crowding in living cells. *Nature Methods*. 12(3):227-229, doi: 10.1038/nmeth.3257, https://dx.doi.org/10.1038/nmeth.3257.

275. Nasir, I., P. L. Onuchic, S. R. Labra, and A. A. Deniz. 2019. Single-molecule fluorescence studies of intrinsically disordered proteins and liquid phase separation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 1867(10):980-987, doi: 10.1016/j.bbapap.2019.04.007, https://dx.doi.org/10.1016/j.bbapap.2019.04.007.

276. Hu, X., L. Hong, M. Dean Smith, T. Neusius, X. Cheng, and Jeremy. 2016. The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time. *Nature Physics*. 12(2):171-174, doi: 10.1038/nphys3553, https://dx.doi.org/10.1038/nphys3553.

277. Guo, S.-M., R. Veneziano, S. Gordonov, L. Li, E. Danielson, K. Perez De Arce, D. Park, A. B. Kulesa, E.-C. Wamhoff, P. C. Blainey, E. S. Boyden, J. R. Cottrell, and M. Bathe. 2019. Multiplexed and high-throughput neuronal fluorescence imaging with diffusible probes. *Nature Communications*. 10(1), doi: 10.1038/s41467-019-12372-6, https://dx.doi.org/10.1038/s41467-019-12372-6.

278. Sabari, B. R., A. Dall'Agnese, and R. A. Young. 2020. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci*. 45(11):961-977, doi: 10.1016/j.tibs.2020.06.007.

279. Strom, A., and C. Brangwynne. 2019. The liquid nucleome – phase transitions in the nucleus at a glance. *Cell Science*. 132(22), doi: 10.1242/jcs.235093, https://dx.doi.org/10.1242/jcs.235093.

280. Debye, P., and L. Pauling. 1925. The inter-ionic attraction theory of ionized solutes. IV. The influence of variation of dielectric constant on the limiting law

for small concentrations. *Journal of the American Chemical Society*. 47(8):2129-2134.

281. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 14(1):33-38, doi: https://doi.org/10.1016/0263-7855(96)00018-5, https://www.sciencedirect.com/science/article/pii/0263785596000185.

282. Giorgino, T. 2014. Computing 1-D atomic densities in macromolecular simulations: The density profile tool for VMD. *Computer Physics Communications*. 185(1):317-322, doi: 10.1016/j.cpc.2013.08.022.

283. Viswanathan, G. M., M. A. G. Portillo, E. P. Raposo, and M. G. E. da Luz. 2022. What Does It Take to Solve the 3D Ising Model? Minimal Necessary Conditions for a Valid Solution. *Entropy (Basel)*. 24(11), doi: 10.3390/e24111665.

284. Semenyuk, P., and V. Muronetz. 2019. Protein Interaction with Charged Macromolecules: From Model Polymers to Unfolded Proteins and Post-Translational Modifications. *Int J Mol Sci*. 20(5), doi: 10.3390/ijms20051252.

285. Zhao, Y., R. Cortes-Huerto, K. Kremer, and J. F. Rudzinski. 2020. Investigating the Conformational Ensembles of Intrinsically Disordered Proteins with a Simple Physics-Based Model. *J Phys Chem B*. 124(20):4097-4113, doi: 10.1021/acs.jpcb.0c01949.

286. Gong, X., Y. Zhang, and J. Chen. 2021. Advanced Sampling Methods for Multiscale Simulation of Disordered Proteins and Dynamic Interactions. *Biomolecules*. 11(10), doi: 10.3390/biom11101416.