Max Planck Graduate Center
mit der Johannes Gutenberg-Universität

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

MAX PLANCK GESELLSCHAFT

Structural and Dynamic Characteristics of Water in Biological Systems

Dissertation

for the award of the academic degree of

"Doctor rerum naturalium" (Dr. rer. nat.)

of the faculties:

08 - Physics, Mathematics and Computer Science

09 - Chemistry, Pharmacy, Geography and Earth Sciences

10 - Biology

University Medicine

submitted by

Daniel Chavez Rojas

Mainz, 2025

Supervisor: Prof. [redacted]

Co-supervisor: Prof. [redacted]

Direct supervisor: Dr. [redacted]

Co-direct supervisor: Dr. [redacted]

Table of Contents

Zusammenfassung **VI**

Summary **X**

1 Introduction **1**

1.1 Background and motivation 5

1.1.1 Protein structure 5

1.1.2 Relevance of protein-protein, protein-solvent, lipid-solvent and protein-lipid interactions 7

1.1.3 Lipid-solvent interactions 9

1.1.4 Relevance to biological systems and related pathologies 12

1.1.5 Motivation behind studying systems that are biological relevant and experimentally 12

1.2 Research objectives 13

1.2.1 Scope of this dissertation 13

1.2.2 Hypothesis and research questions 13

1.2.3 Significance of research contributions 14

2 Theoretical background **17**

2.1 Statistical mechanics 18

2.1.1 Microstates and Macrostates 18

2.1.2 Thermodynamical ensembles and probability distribution 18

2.1.3 Entropy 23

2.2 Solvation 25

2.2.1 Solvation Energy 25

2.2.2 Factors Influencing Solvation Energy 26

2.2.3 Phase separation theory 26

2.2.4 Phase separation of polymers 28

2.2.5 Polymer theory for biological systems 33

2.3 Computational methods 34

2.3.1 Molecular dynamics simulations . . . 34
2.3.2 Thermodynamic Variables . . . 35
2.3.3 Dynamics algorithms . . . 42
2.3.4 Coarse-grained models and backmapping schemes . . . 44
2.4 Graph analysis formalism . . . 46
2.5 Spectroscopy techniques . . . 50
2.5.1 Two dimensional infrared spectroscopy . . . 50
2.5.2 Sum frequency generation spectroscopy . . . 52

3 Methods and results **53**
3.1 Water orientation around lipid monolayers . . . 53
3.1.1 Simulations . . . 54
3.2 Water structure and dynamics around protein condensates . . . 61
3.2.1 Slowing of water dynamics inside protein condensates . . . 61
3.2.2 Water structure around FUS LC protein . . . 76
3.2.3 Water structure in protein condensate microenvironments . . . 78

4 Discussion **83**
4.1 Lipid Organization and Water orientation . . . 83
4.2 Slowing of water dynamics inside FUS LC protein condensate . . . 84
4.3 Water structure around FUS LC protein . . . 85

5 Conclusion **88**
5.1 Biological implications . . . 88
5.2 Future research directions . . . 90
5.2.1 Lipid organization and water orientation . . . 90
5.2.2 Water structure and dynamics around protein condensates . . . 91

Acknowledgements **93**

Bibliography **94**

Publication statement **103**

Permission rights **105**

List of Abbreviations

PC phosphatidylcholine

PE phosphatidylethanolamine

GPCR g protein-coupled receptors

AML acute myeloid leukemia

MD molecular dynamics

CG coarse-grained

2D-IR two-dimensional infrared

IR infrared

SFG sum frequency generation

BCs biomolecular condensates

LLPS liquid-liquid phase separation

FFCs frequency–frequency correlations

FUS fused in sarcoma

CLS center line slope

List of Figures

1.1 Protein Hierarchical organization 6
1.2 Protein-water hydrogen bond 8

2.1 Phase separation diagram 32
2.2 Graph analysis 46

3.1 Lipid chemistry 54
3.2 lipid monolayers simulations 55
3.3 Tetrahedral order parameter 56
3.4 Histograms of tetrahedral order parameter at lipid interface PC 57
3.5 Histograms of tetrahedral order parameter at lipid interface PE 57
3.6 Depiction of the vectors used as reference for calculating orientation of water molecules in reference to the membrane 58
3.7 Water angle vs tail type 59
3.8 Depiction of lipid tilt angle 60
3.9 Tilt angle histogram 60
3.10 Experimental samples of FUS LC BC vs diluted phase 62
3.11 Linear expectra of FUS LC samples 63
3.12 Infrared spectra from MD simulations 65
3.13 2D IR spectra of FUS LC 67
3.14 CLS lines for all samples 69
3.15 hydrogen bond criteria 74
3.16 FUS clasification and time autocorrelation plot 75
3.17 Tetrahedral order parameter distribution 77
3.18 Contact map 79
3.19 Histograms of tetrahedral order parameter distribution 80
3.19 Histograms of tetrahedral order parameter distribution 81
3.20 Density map plot of aminoacid degree 81
3.20 Density map plot of aminoacid degree (continuation) 82

Zusammenfassung

Wasser ist für das Leben unentbehrlich und spielt eine wesentliche Rolle in biologischen Systemen, von molekularen Interaktionen bis hin zu physiologischen Prozessen im großen Maßstab. Wassermoleküle bestehen aus zwei Wasserstoffatomen und einem Sauerstoffatom (H_2O) und haben eine gebogene Struktur, die ein polares Molekül ergibt. Diese Struktur ermöglicht die Bildung von Wasserstoffbrücken, sowohl zwischen Wassermolekülen als auch mit anderen polaren Substanzen, was sie zu hervorragenden Lösungsmitteln macht. Darüber hinaus ist die Untersuchung von Wasser von grundlegender Bedeutung für das Verständnis der biologischen Welt. In der Molekularbiologie fungiert Wasser nicht nur als Medium für den Nährstofftransport, sondern hat auch direkte Auswirkungen auf die Faltung, Stabilität und Funktion von Proteinen sowie auf die Dynamik von Lipidmembranen.

Insbesondere interagieren Wassermoleküle eng mit Lipidkopfgruppen und bilden Hydratationsschichten, die die Lipiddynamik, Diffusion und Konformation erheblich beeinflussen. Die Untersuchung dieser Wechselwirkungen kann das Design und die Wirksamkeit von Arzneimittelabgabesystemen wie Liposomen beeinflussen oder bei der Entwicklung völlig neuer Abgabesysteme helfen, die die Bioverfügbarkeit erhöhen oder auf bestimmte Gewebe abzielen.

Wasser spielt auch bei Proteinen eine entscheidende Rolle, indem es deren Struktur durch Wechselwirkungen mit polaren und geladenen Gruppen auf der Oberfläche des Proteins stabilisiert. Wasser fungiert auch als Medium zur Erleichterung biochemischer Wechselwirkungen und ermöglicht es Proteinen, mit anderen Molekülen zu interagieren. Wassermoleküle stehen in enger Wechselwirkung mit Proteinen, indem sie Hydratationsschalen um deren Oberflächen bilden. Diese Hydratationshüllen beeinflussen die Proteindynamik, d. h. wie sich Proteine bewegen, an andere Moleküle binden und biochemische Reaktionen durchführen. Die Untersuchung dieser Wechselwirkungen kann erhebliche Auswirkungen auf die Entwicklung von Arzneimitteln haben, da sie hilft zu

verstehen, wie sich Proteine in verschiedenen Umgebungen verhalten. Dieses Wissen kann die Entwicklung von Therapien auf Proteinbasis leiten oder die Verabreichung von Arzneimitteln, die auf bestimmte Proteine abzielen, verbessern, wodurch die Wirksamkeit verbessert und die Nebenwirkungen verringert werden können.

In dieser Dissertation konzentrieren wir uns hauptsächlich auf die Charakterisierung von Wasser im Zusammenhang mit Lipiden und Proteinkondensatsystemen. Zunächst untersuchen wir Lipid-Wasser-Grenzflächen, um zu verstehen, wie Unterschiede in der Lipidchemie die Wasserorientierung beeinflussen können. Zu diesem Zweck haben wir Molekulardynamiksimulationen (MD) von Membranen durchgeführt, die aus verschiedenen Lipidtypen bestehen, und die Ausrichtung der Wassermoleküle an ihrer Grenzfläche analysiert. Unsere Arbeit zeigt, dass die Schwanzchemie der Lipide eine geringere Rolle bei der Ausrichtung der Wassermoleküle spielt als die Lipidorganisation. Dies zeigt, wie die räumliche Anordnung der Lipide und das kollektive Verhalten die Wasserstruktur an der Lipidgrenzfläche beeinflussen.

Ein weiteres biologisches System, das wir untersuchen, ist das Fused-in-Sarcoma-Protein (FUS). Dieses Protein wurde aufgrund seiner Neigung zur Selbstorganisation in Stressgranula innerhalb von Zellen und der Bildung von Einschlusskörpern, die bei amyotropher Lateralsklerose und anderen neurodegenerativen Krankheiten beobachtet werden, eingehend untersucht. Die Hauptursache für diese Selbstorganisation ist die intrinsisch ungeordnete LC-Domäne von FUS. Verschiedene Faktoren wie pH-Wert, Temperatur, Ionenstärke und molekulare Verdrängung können die LC-Kondensation von FUS auslösen, was es zu einem idealen Modell für die Untersuchung von Proteinkondensaten macht.

Proteinkondensate schaffen unterschiedliche biochemische Umgebungen, die spezifische Reaktionen fördern oder hemmen können, ohne dass membrangebundene Organellen erforderlich sind. Durch die Konzentration von Enzymen und Substraten erhöhen sie die Reaktionsgeschwindigkeit, während sie durch den Ausschluss anderer Moleküle unerwünschte Wechselwirkungen unterdrücken können. Diese Form der Kompartimentierung ist für Prozesse wie den RNA-Stoffwechsel entscheidend.

In unserer Arbeit haben wir die Wasserstoffbrückenbindungen in diesen Kondensaten untersucht, die ein entscheidender Faktor bei der Solvatisierung sind. Wir haben uns auf die Erforschung der komplizierten Wasser-Protein-Wasserstoffbrückenbindungsdynamik innerhalb der Kondensate konzentriert, die von der niedrig komplexen Domäne von FUS LC gebildet werden. Um dies zu erreichen, haben wir ein multidisziplinäres und interinstitutionelles Forschungsprojekt durchgeführt. Dies führte zu einer Zusammenarbeit zwischen dem Max-Planck-Institut für Polymerforschung, insbesondere den Abteilungen Theorie und Molekülspektroskopie, und dem Department of Biomedical Engineering der University of Texas, Austin. In dieser Zusammenarbeit analysierten wir die Dynamik von Wasser in FUS-LC-Kondensaten und verglichen sie mit der Dynamik von Wasser in vollständig solvatisiertem FUS-LC-Protein, sowohl experimentell als auch theoretisch. Sayantan Chatterjee stellte die FUS-LC-Proben zur Verfügung, [redacted] führte 2D-Infrarot-Spektroskopie (2D IR) durch, um die FUS-LC-Spektren aufzuzeichnen, und ich führte die Molekulardynamiksimulationen und die theoretische Analyse durch. In Kombination liefern Spektroskopie und MD-Simulationen ein klares Bild des molekularen Verhaltens.

Die 2D-IR-Experimente zeigten, dass die spektrale Diffusionsdynamik von FUS LC in kondensierten Tröpfchen etwa doppelt so langsam ist wie die Dynamik in einer homogenen Lösung. Dies veranlasste uns zu der Hypothese, dass die beobachteten Unterschiede in der Dynamik auf eine Verlangsamung der Hydratationsdynamik für FUS LC innerhalb des Kondensats zurückzuführen sind. Um diese Hypothese zu untersuchen, führten wir vollständig atomistische MD-Simulationen von FUS LC und seiner solvatisierten Form durch. Wir analysierten die Lebensdauer der Wasserstoffbrückenbindungen zwischen den Amidgruppen (C=O) der Proteinketten und den Wassermolekülen sowohl in der kondensierten als auch in der solvatisierten Phase, indem wir die Zeit-Autokorrelationsfunktion, auch bekannt als Überlebensfunktion, berechneten. Diese Analyse bestätigte, dass die Lebensdauer der Wasserstoffbrückenbindungen zwischen Wasser und den Amidgruppen bei FUS LC in den Kondensaten deutlich länger ist (etwa 40%) als bei verdünnter FUS LC. Unsere Ergebnisse zeigen, dass die Wasserdynamik in den Kondensaten langsamer ist als in den verdünnten Gegenstücken. Wir vermuten, dass diese unterschiedliche Dynamik

auf Pikosekunden-Zeitskalen auf eine verlangsamte Wasserstoffbrückenbindungsdynamik in der das Kondensat umgebenden Hydratationsschale zurückgeführt werden kann.

Unter Ausnutzung der detaillierten atomistischen Einblicke, die MD-Simulationen in die Struktur von Wassermolekülen und Proteinen innerhalb des Kondensats bieten, erweitern wir unsere Arbeit, um zu untersuchen, wie sich Wassermolekülnetzwerke innerhalb des Kondensats anordnen. Aufgrund der riesigen Datenmenge von mehr als 3 Milliarden Datenpunkten haben wir einen auf Graphenanalyse basierenden Formalismus entwickelt, um die Protein-Protein- und Protein-Wasser-Wechselwirkungen mit einer Auflösung auf Aminosäureebene zu untersuchen und zu klassifizieren, wobei wir festgestellt haben, dass Tyrosin und Glutamin im Verhältnis zu anderen Aminosäuren Mikroumgebungen mit dicht gedrängten Kontaktnetzwerken bilden. Um die Wasserstruktur in diesen Mikroumgebungen zu charakterisieren, berechneten wir anschließend die tetraedrischen Ordnungsparameter, die eine Störung der Wasserstruktur in überfüllten Umgebungen zeigten. Dank der Kombination dieser leistungsstarken Techniken, der Graphenanalyse und der MD-Simulationen konnten wir die Hauptakteure bei der Bildung komplizierter Protein-Protein- und Protein-Wasser-Netzwerke aufklären.

Zusammenfassend lässt sich sagen, dass ein so einfaches Molekül wie Wasser nicht nur ein Lösungsmittel ist, sondern eine aktive Komponente biologischer Systeme, die die Wechselwirkungen in Lipiden und Proteinkondensaten gestaltet. Mit Hilfe der atomistischen Auflösung, die Molekulardynamiksimulationen bieten, konnten wir die feinen Details der Wasserdynamik im Kontext dieser biologischen Systeme erfassen.

Summary

Water is indispensable to life, playing an intrincate role in biological systems, from molecular interactions to large scale physiological processes. Water molecules are composed of two hydrogen atoms and one oxygen atom (H_2O) and have a bent structure creating a polar molecule. This structure allows for the formation of hydrogen bonds, both between water molecules and with other polar substances. which makes them excellent solvents. Moreover, the study of water is fundamental for understanding the biological world. In the case of molecular biology, besides acting as a medium for nutrient transport, water interactions have a direct effect on protein folding, stability, and function as well as lipid membrane dynamics.

Specifically, water molecules interact closely with lipid head groups, forming hydration layers that significantly affect lipid dynamics, diffusion, and conformation. Studying these interactions can influence the design and efficacy of drug delivery systems, such as liposomes, or aid in the development of completely new delivery systems that increase bioavailability or target specific tissues.

Water also plays a crucial role with proteins by stabilizing their structure through interactions with polar and charged groups on the protein's surface. Water also acts as a medium for facilitating biochemical interactions, allowing proteins to interact with other molecules. Water molecules interact closely with proteins by forming hydration shells around their surfaces. These hydration shells influence protein dynamics, including how proteins move, bind to other molecules, and carry out biochemical reactions. Studying these interactions can have significant implications for drug design, as it helps in understanding how proteins behave in various environments. This knowledge can guide the development of protein-based therapies or enhance the delivery of drugs that target specific proteins, potentially improving efficacy and reducing side effects.

In this dissertation, we primarily focus on characterizing water in the context of lipids and protein condensate systems. Firstly, we explore lipid-water interfaces aiming to understand how differences in lipid chemistry can affect water orientation. To do this, we performed molecular dynamics (MD) simulations of membranes composed of different lipid types, and analyzed the orientation of water molecules at their interface. Our work shows that lipid tail chemistry plays a lesser role in affecting the orientation of water molecules, compared to lipid organization. This highlights how lipid spatial arrangement and collective behavior impact water structure at the lipid interface.

Another biological system that we study is the Fused in sarcoma (FUS) protein. This protein has been extensively studied due to its propensity for self-assembly into stress granules within cells, and the formation of inclusion bodies observed in amyotrophic lateral sclerosis and other neurodegenerative diseases. The key driver behind this self-assembly is the intrinsically disordered low-complexity (LC) domain of FUS. Various factors, including pH, temperature, ionic strength, and molecular crowding, can trigger FUS LC condensation, making it an ideal model for investigating protein condensates.

Protein condensates create distinct biochemical environments that can promote or inhibit specific reactions without the need for membrane bound organelles. By concentrating enzymes and substrates, they enhance reaction rates, whereas their exclusion of other molecules can suppress undesired interactions. This form of compartmentalization is critical in processes like RNA metabolism.

In our work, we shed light into the hydrogen-bonding environment within these condensates, a crucial factor in solvation. We focused on the exploration of the the intricate water-protein hydrogen-bond dynamics within condensates formed by the low complexity domain of FUS LC. To achieve this, we conducted a multidisciplinary and inter-institutional research project. This resulted in the collaboration between the Max Planck Institute of Polymer Research, specifically its Theory and Molecular Spectroscopy departments, and the Department of Biomedical Engineering at the University of Texas, Austin. In this collaborative effort, we analyzed the dynamics of water within FUS LC condensates and contrasted it with the dynamics of water in fully solvated FUS LC protein, both experimentally and theoretically. Specifically, Sayantan Chatterjee provided the

FUS LC samples, [redacted] conducted 2D infrared (2D IR) spectroscopy to record the FUS LC spectra, and I performed the molecular dynamics simulations and theoretical analysis. When used together, Spectroscopy and MD simulations provide a clear picture of molecular behavior.

The 2D IR experiments indicated that the spectral diffusion dynamics of FUS LC in condensed droplets is approximately twice as slow compared to its dynamics in a homogeneous solution. This led us to hypothesize that the observed differences in dynamics result from a deceleration of hydration dynamics for FUS LC within the condensate. To investigate this hypothesis, we conducted fully atomistic MD simulations of FUS LC and its solvated form. We analyzed the lifetimes of hydrogen bonds between the amide (C=O) groups of the protein chains and water molecules in both the condensed and solvated phases by calculating the time autocorrelation function, also known as survival function. This analysis confirmed that hydrogen bond lifetimes between water and the amide groups are significantly longer (approximately 40%) for FUS LC in the condensates, compared to dilute FUS LC. Our findings reveal that water dynamics inside the condensate are slower than their dilluted counterparts. We suggest that these distinct dynamics on picosecond time scales can be attributed to slowed hydrogen-bonding dynamics in the hydration shell surrounding the condensate.

Taking advantage of the detailed atomistic level insights that MD simulations provide into the structure of water molecules and proteins within the condensate, we extend our work to investigate how water molecule networks arrange themselves inside the condensate. Due to the huge amount of data, consisting of more than 3 billion data points, we designed a graph analysis based formalism to study and classify the protein-protein and protein-water interactions at an amino acid level resolution, where we observed that tyrocine and glutamine, form microenvironments of crowded contact networks in relation to other amino acids. Then, in order to characterize the water structure in these microenvironments, we calculated the tetrahedral order parameters, showing a disruption of water structure in crowded environments. Thanks to the combination of such powerful techniques, graph analysis and MD simulations allowed us to elucidate the key players in forming intricate protein-protein and protein-water networks.

In summary, we showcase that a molecule as simple as water, is not merely a solvent, but an active component of biological systems, shaping interactions in lipids as well as in protein condensates. Aided by the atomistic resolution that molecular dynamics simulations provide, we were able to capture the fine details of how water dynamics function in the context of these biological systems.

Chapter 1

Introduction

Proteins are basic structural, functional, and regulatory molecules present in all living organisms. They are encoded in the genetic code and synthesized through a process called translation, which occurs in ribosomes. Translation is the mechanism by which the sequence of a messenger RNA molecule is decoded to produce a specific sequence of amino acids, forming a protein. Proteins are made from basic building blocks called amino acids and are involved in various biological processes, making them essential for cellular function. For this reason, their study is fundamental to understanding the macromolecular world.(*1*).

The study of proteins has a history spanning centuries of scientific research. The term "protein" itself has a Greek origin, derived from the word "proteios", meaning primary or first place. However, it wasn't until the 19th century that scientists recognized proteins as distinct entities within living organisms, separate from other biological macromolecules such as carbohydrates and lipids, In the early 18th century, French chemist Hilaire Rouelle coined the term "albumen" to describe egg white, marking one of the earliest identifications of a protein(*2*). In 1838, Swedish chemist Jöns Jakob Berzelius introduced the term "protein" to encompass a class of nitrogen-rich compounds in living tissues.

Further advancements in protein chemistry occurred in the late 19th and early 20th centuries. Friedrich Miescher's discovery of nucleic acids briefly shifted focus, but attention soon returned to proteins as the isolation of various types of proteins, such as enzymes, antibodies, and hemoglobin, provided diverse research paths.

The 20th century brought the introduction of X-ray crystallography, a groundbreaking technique for determining the three-dimensional structures of structured proteins. This method works by directing a beam of X-rays at a crystallized protein sample. As the X-rays pass through the crystal, they are scattered by the electrons in the atoms of the protein, creating a diffraction pattern. This pattern, when captured on a detector, contains information about the atomic arrangement within the crystal. With this information is possible to construct an electron density map, which reveals the precise positions of atoms in structured proteins. In 1957, Francis Crick proposed the central dogma of molecular biology (*3*), suggesting an unidirectional flow of information from DNA to RNA to proteins. Which became a guiding principle for molecular biology research.

The late 20th and early 21st century witnessed the rise of proteomics, a field dedicated to the large-scale study of proteins, their structures, functions, and interactions within a biological system.

With innovations in mass spectrometry, protein sequencing, cryo electron microscopy and recombinant DNA technology. In recent years, systems biology has gained traction, considering proteins within the context of large biological ensembles. This integrative perspective aims to understand the interplay between proteins and other biomolecules.

Nevertheless, as protein structures were being deciphered, the limitations of structural information obtained from methods like X-ray crystallography, which provide a snapshot of a protein in a single, fixed conformation became apparent. While these methods are invaluable for understanding protein structure, they do not capture the inherent flexibility and dynamic behavior of proteins in their natural, fluctuating states.

Proteins are not rigid entities; they undergo constant movements, such as folding, conformational changes, and interactions with other molecules, all of which are essential to their function. This highlighted the need for methods that capture the dynamic nature of proteins. This led to the use of molecular dynamics, whose origins can be traced back to the 1950s and 1960s. In the 1970s, McCammon and Gelin (*4*) conducted the first molecular dynamics simulation of a protein.

Over subsequent decades, molecular dynamics expanded with advancements in computational power and methodology. Supercomputers enabled more extensive and realistic simulations trough the use of force fields specifically designed to capture the movement of biological molecules, capturing the movements of thousands of atoms over longer timescales. A force field is a mathematical model used to describe the interactions between atoms and molecules through potential energy functions that calculate the forces acting on particles. Force fields, such as CHARMM(*5*) and AMBER(*6*), were developed to describe atom interactions, more the accurately. These force fields are based on classical approximations, meaning they use classical mechanics rather than quantum mechanics to approximate atomic and molecular behavior. This approach enhances the accuracy of simulations while maintaining computational feasibility.

Molecular dynamics is playing a vital role in understanding basic principles of protein folding, providing insights into the processes by which a linear chain of amino acids can assume its functional three-dimensional structure. Currently simulations can provide guidance into possible folding trajectories, driving forces, and key intermediates, and can shed light on protein unfolding dynamics, revealing the stability of different structural motifs and the impact of external factors.

Water molecules have a unique ability to form intricate networks of hydrogen bonds due to their polar nature, which allows them to interact with a variety of functional groups present in proteins and lipids. Around proteins, water molecules form hydration shells that envelop the protein surface, contributing significantly to the protein's stability and structural dynamics (*7*). These hydration shells not only solvate the protein,

but also play a vital role in mediating protein-protein interactions, ligand binding, and enzymatic activity.

Understanding the interplay between water molecules and biological molecules like proteins and lipids is crucial in comprehending the fundamental mechanisms governing biological processes. Their interactions play a pivotal role in dictating the structure, stability, and function of macromolecules (*7*).

Water molecules have a unique ability to form intricate networks of hydrogen bonds due to their polar nature, which allows them to interact with a variety of functional groups present in proteins and lipids. Around proteins, water molecules form hydration shells that envelop the protein surface, contributing significantly to the protein's stability and structural dynamics (*7*). These hydration shells not only solvate the protein, but also play a vital role in mediating protein-protein interactions, ligand binding, and enzymatic activity.

The mobility of water molecules around proteins and lipids are highly dynamic and heterogeneous, for these reason their role at the interface of proteins and lipids are influenced by factors such as temperature (*8*), pressure (*9*), pH (*10*), and the specific chemical and structural features of the biomolecules involved. Molecular dynamics simulations, spectroscopic techniques like NMR (*11*), X-ray crystallography (*12*), and theoretical models (*13*) have provided insights into the spatiotemporal behavior of water molecules in these biological systems

A comprehensive understanding of water structure and dynamics around proteins and lipid systems is essential for unraveling the underlying principles governing biological functions, drug interactions, and the design of novel therapeutics. Investigating these interactions at the molecular level not only sheds light on fundamental biological processes but also holds immense potential for various biotechnological and pharmaceutical applications. The exploration of water behavior within these biological contexts

continues to evolve, offering opportunities for further research and discoveries in the realm of physics, biophysics, biochemistry, and structural biology.

1.1 Background and motivation

1.1.1 Protein structure

The functionality of proteins is not solely determined by their individual structures but is also influenced by their interactions with other bio molecules. Among these interactions, protein-protein, protein-solvent, and protein-lipid interactions stand out as pivotal contributors to the dynamic and intricate nature of cellular processes.

The structure of proteins is hierarchically organized. At the primary level, proteins are linear chains of amino acids linked together by peptide bonds. The sequence of amino acids in a protein is determined by the genetic code in an organism's DNA. This sequence, known as the protein's primary structure (Figure 1.1), serves as a foundation for its three-dimensional shape and function (*14*).

The secondary structure of proteins arises from folding and interaction of neighboring amino acids. Common secondary structures include alpha helices and beta sheets, which result from hydrogen bonding between amino acids. These structures contribute to the overall stability and function of the protein (*14*). Water, with its ability to mediate hydrogen bonding, plays a vital role in ensuring the proper formation of these secondary structures, contributing significantly to protein stability and function.

Tertiary structure refers to the three-dimensional arrangement of the entire polypeptide chain. It is influenced by various interactions, such as hydrogen bonding, disulfide bridges, ionic bonds, and hydrophobic interactions. The final, functional form of a protein is often determined by its tertiary structure (*14*).

In some cases, proteins may consist of multiple polypeptide chains, each with its own tertiary structure. The quaternary structure refers to the arrangement of these individual chains and the interactions between them. Hemoglobin, for instance, is a globular protein with quaternary structure, composed of four polypeptide subunits (*14*).

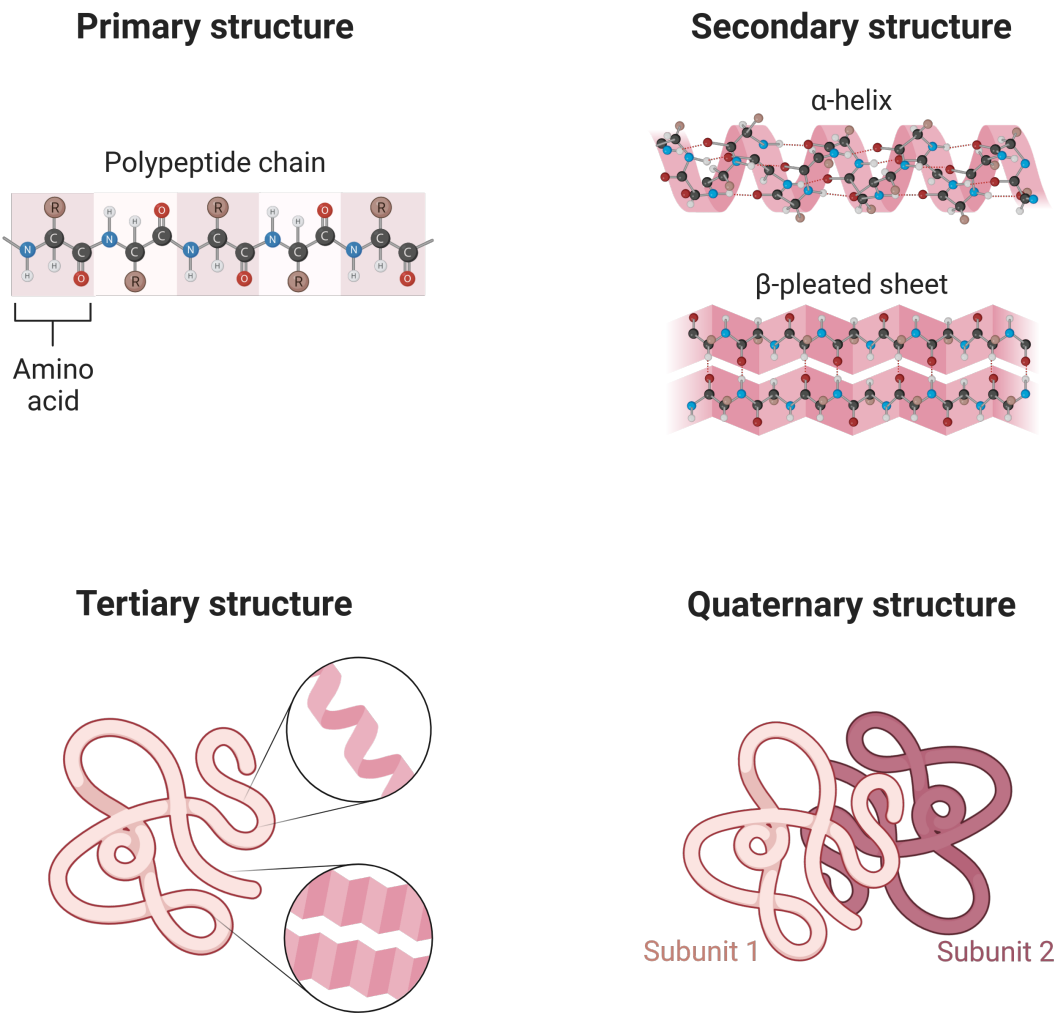

Figure 1.1: **Protein Hierarchical organization.** Proteins have four levels of hierarchical organization, consisting of primary, secondary, tertiary and quaternary structures (Created with BioRender.com).

Proteins can have different functions. They can serve as enzymes, catalysts that facilitate biochemical reactions, as structural proteins providing shape to cells and tissues, as transport aiding in the movement of substances across membranes, as

hormones, acting as signaling molecules, or as antibodies, contributing to the immune system's defense against pathogens(*14*).

Proteins that only have a primary structure are known as intrinsically disordered proteins, or IDPs. In some cases proteins may contain structured and unstructured regions; in this case, unstructured sections are called intrinsically disordered regions or IDRs.

1.1.2 Relevance of protein-protein, protein-solvent, lipid-solvent and protein-lipid interactions

Protein-protein interactions

Protein-protein interactions orchestrate the regulation of numerous cellular processes, including signal transduction, gene expression, and metabolic pathways. Understanding the specificity, affinity, and dynamics of these interactions provides insights into the coordination and fine-tuning of cellular functions (*15*). Likewise, deregulation of protein-protein interactions is implicated in various diseases, such as cancer, neurodegenerative disorders, and infectious diseases (*16*).

However, protein-protein interactions play an important role in drug discovery and design. Many pharmaceutical molecules target protein-protein interactions to modulate cellular pathways. A comprehensive understanding of these interactions facilitates the rational design of drugs that selectively interfere with specific protein complexes (*17*) (*18*).

Protein-solvent interactions

The solvent environment significantly influences protein stability, aggregation, and folding. An example of solvent affecting protein behavior is that the solubility of a protein can change with the pH of the solvent. For instance, casein, a protein found in

milk, is soluble in the pH range of about 6.6 to 7.0 (*19*). However, if the pH is lowered, casein can precipitate out of solution due to changes in the ionization of its amino acid side chains. This phenomenon is often used in cheese-making, where the acidic environment leads to the coagulation of casein. Another example of solvent affecting protein behavior is how the presence of salts in the solvent can also affect protein solubility. For example, high ionic strength can shield the charges on a protein's surface. This can lead to reduced electrostatic repulsion between protein molecules, which may lead to protein aggregation or precipitation (*20*). At low ionic strengths, proteins may be more soluble due to the increased repulsion between their similarly charged surfaces. Furthermore, hydrogen bonding between the protein surface and water molecules (1.2) in the solvent plays a critical role in maintaining protein solubility and stability. These bonds help stabilize the protein's structure by facilitating hydration of the polar and charged groups on the protein surface. For this reason is necessary to understand the protein-solvent interactions which allows to elucidate mechanisms for proper protein function (*7*).

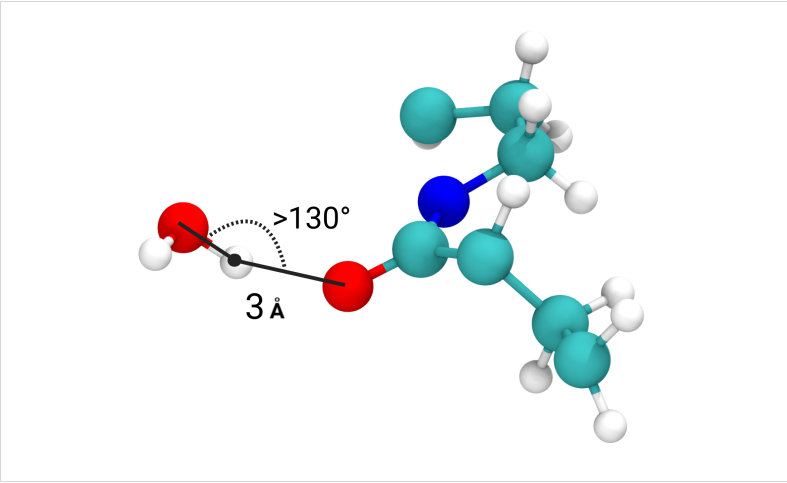

Figure 1.2: **Protein-water hydrogen bond.** Protein backbone can form hydrogen bonds with water molecules, the figure shows a water molecule donating it's hydrogen atom to the oxygen of the amide group of an amino acid in order to form an hydrogen bond

Another biological process where solvent plays a fundamental role is enzyme catalysis one example where solvent affects protein behavior is with serine proteases these are a class of enzymes that utilize a serine residue in their active site to perform hydrolysis

of peptide bonds. In an aqueous environment, water serves as a medium for substrate binding and product release, and it also participates in the reaction by stabilizing the transition state and providing protons during hydrolysis. But when instead of water the protein is surrounded by an organic solvent like ethanol, it can alter the polarity and dielectric constant of the medium, affecting protein conformation and dynamics, and in this case enhance the catalytic activity of some serine proteases by promoting a more flexible enzyme structure. However, at very high concentrations, organic solvents can denature the enzyme, reducing its activity. Another reason by which organic solvent can affect catalysis is due the solvent polarity, this can influence the ionization states of amino acids in the active site, impacting the enzyme ability to interact with substrates (*20*). these are some examples of how solvent molecules participate in enzyme catalysis by influencing the micro-environment around active sites. Studying the intricacies of these protein-solvent interactions can aid to understand enzyme kinetics and substrate specificity (*21*).

1.1.3 Lipid-solvent interactions

Lipids can be broadly defined as amphiphilic small molecules, the amphiphilic nature of lipids allows them to form structures such as vesicles and membranes in an aqueous environment (*22*). Phospholipids are a particular class of lipids, these molecules are caracterized by having an hydrophilic head containing a phosphate group, and most of them contain two hydrophobic tails composed of fatty acids (*23*).

Lipid headgroup chemistry

Of particular interest for this dissertation are two particular types of lipid headgroups. Phosphatidylcholine (PC) and phosphatidylethanolamine (PE) which are two major classes of phospholipids that differ by the chemistry of their headgroup. PC lipids have a choline group attached to their phosphate heads, which contains a trimethylamine structure. This headgroup is bulky and more hydrophilic than PE due to the presence of charged atoms (*24*). For this reason, PC lipids tend to pack more loosely.

This loose packing allows for higher hydration around the lipid headgroups, as water can easily penetrate and interact with the exposed phosphate and choline moieties (*24*).

PE lipids, on the other hand, contain an ethanolamine headgroup, which is small and has a primary amine (-NH2) attached to an ethyl group. This headgroup has a more compact and less bulky structure compared to PC (*24*). The smaller ethanolamine headgroup in PE lipids allows for tighter packing. This tight packing reduces the amount of water that can interact with the headgroup, creating a more hydrophobic environment around lipid-water interfaces.

Lipid tail saturation

Saturation refers to whether the fatty acid chains have double bonds (unsaturated) or only single bonds (saturated) along their lipid tails, each impacting the lipid's behavior and way to interact with with water.

Saturated fatty acid chains contain no double bonds, meaning that the hydrocarbon chain is fully saturated with hydrogen atoms. As a result, the tails are straight and can pack closely together. Tighter packing reduces the hydration shell around the lipids, due to the lack of space for water molecules to interact with the headgroups or to penetrate into the membrane. This less hydrated surface and can lead to a more structured, but thinner layer of water molecules around the lipid membrane (*25*).

On the other hand, unsaturated fatty acid chains have one or more double bonds, introducing kinks in the tail. These kinks prevent the tails from packing as tightly, creating more space between the lipids, resulting in a more fluid and dynamic membrane. This conformation increases the hydration of the membrane and enhances the interaction between water and the lipid headgroups.

Lipid tail

Interactions between short lipid tails experience reduced Van der Waals forces, resulting in a fluid membrane with enhanced water interactions at the surface. This leads to a more dynamic and fluctuating hydration shell.

Long tails, however, increase Van der Waals interactions, resulting in more rigid and ordered membranes. This rigidity reduces the mobility of the headgroups, decreasing their interaction with surrounding water molecules. In this conformation, the water molecules remain largely restricted to interact with the polar headgroups (*26*).

In summary, longer and fully saturated lipids will form a more structured layer than shorter unsaturated lipid tails.

Protein-lipid interactions

Proteins interact with lipids in cell membranes, influencing membrane structure, fluidity, and permeability. This is critical for cellular processes such as signal transduction, membrane trafficking, and cell-cell communication (*27*). Exploiting protein-lipid interactions is integral to the design of drug delivery systems. Tailoring these interactions enables the development of carriers that enhance drug solubility, stability, and targeted delivery (*28*). An example of a lipid interacting protein that has taken relevance in recent years are G protein-coupled receptors (GPCRs). GPCRs have a characteristic structure made up of seven transmembrane alpha-helices. These helices span the lipid bilayer of the cell membrane, with portions of the receptor extending outside the cell and a portion extending into the cytoplasm (*29*). When a signaling molecule, such as a hormone protein, binds to the extracellular portion of the GPCR, it induces a conformational change in the receptor. This conformational change activates an associated G-protein on the intracellular side. The G-protein then dissociates into its subunits, which can interact with other intracellular signaling pathways, leading to a cellular response. The lipid bilayer is crucial for anchoring the GPCRs and allowing

the proper alignment and interaction with not only the ligand, but also the G-proteins and downstream signaling molecules. GPCRs are among the most important drug targets, with a significant fraction of therapeutic molecules targeting them. Drugs like beta-blockers (*30*), antihistamines (*31*) and antipsychotics (*32*) exemplify their therapeutic versatility. A recent successful therapeutic case involving a GPCR receptor are the GLP-1 antagonists that have been developed for treating diabetes, and more recently, as weight loss drugs (*33*).

1.1.4 Relevance to biological systems and related pathologies

Water, a seemingly ubiquitous component of biological systems, is more than a simple passive medium for functional macromolecules to diffuse within biological systems. Its unique molecular structure, characterized by hydrogen bonding, shapes the dynamic landscapes of biomolecular interactions. Water molecules dynamically cluster and reorganize, participating in the conformational flexibility of biological molecules.

1.1.5 Motivation behind studying systems that are biological relevant and experimentally

The reason behind investigating systems characterized by both biological relevance and experimental feasibility comes from the importance of establishing a connection between theoretical frameworks and empirical substantiation. This approach serves to amplify the utility of the results, favoring implementation of theoretical understanding and empirical applicability. This dual emphasis enriches the understanding of biological processes. Specifically, our investigation of slowing water dynamics in intrinsically disordered proteins IDPs using spectroscopy and MD simulations exemplifies this approach. Spectroscopy provided direct empirical evidence of altered hydration dynamics near IDPs, while MD simulations offered a detailed theoretical framework to interpret these changes at the molecular level. The integration of these methods not only validated the findings but also provided a deeper understanding of the interaction between IDPs and their hydration shell.

1.2 Research objectives

1.2.1 Scope of this dissertation

In this dissertation, we employ a combined approach of spectroscopy, network science theoretical framework and molecular dynamics simulations to characterize the environmental, structural, and dynamic processes of water molecules within distinct biological systems. Focusing on protein condensates and lipid monolayers.

Molecular dynamics simulations provide an ideal tool for understanding the dynamics of biological systems at an atomistic resolution. The simulations give access to the dynamic and structural properties of water molecules around the disorder protein systems studied in this dissertation.

Network science provides a theoretical formalism for understanding interactions between biomolecular systems and their solvent environments. By abstracting complex structures into network representations, we quantify spatial relationships and connectivity patterns within protein networks and water molecules. Applying this methodology to protein condensates provides insights into the spatial organization and connectivity of water within these condensates.

Integration of these two techniques permits a methodological and efficient path to analyze and interpret the vast amount of data obtained from the simulations.

1.2.2 Hypothesis and research questions

The aim of this dissertation is to answer the following questions:

- How is the orientation of water differ at the interface of lipid monolayers, and does this orientation exhibit sensitivity to variations in lipid chemistry?
- Do the dynamics of water inside protein condensates differ from their diluted counterparts, indicating differences in the interactions within these respective environments?

- What are the defining features of the environment within protein-protein networks?
- Do water molecules exhibit distinct structural arrangements within protein condensates, and if so, are these structural differences localized to specific regions within the condensate environment?

1.2.3 Significance of research contributions

Understanding water structure and orientation at lipid monolayers is crucial for designing drug delivery systems like liposomes and for developing biomimetic materials. The behavior of water at lipid interfaces can affect how these systems encapsulate and release drugs, influencing their efficiency and bioavailability.

In the case of molecular condensates there is evidence linking protein aggregation to the pathogenesis of neurodegenerative diseases. Proteins like FUS, undergo phase separation to form condensates. When this process is dysregulated, it can lead to the formation of solid protein aggregates, mutations in these proteins disrupt their normal phase behavior, promoting pathogenic aggregation and neurotoxicity, which has been linked to neurodegenerative deseases like ALS (*34*).

Another case of pathologies driven by protein condensates are those derived from mistranslocation or mutation. A notable example being oncogenes, like MYC (*35*) and BRD4 (*36*). When mistranslation or mutations occur, these proteins can form condensates (*37*). These condensates play a role in upregulating expression of genes that drive cancer cell proliferation, the study of these condensates has led to the development of inhibitors capable of disrupting these condensates, offering a potential therapeutic target in cancers such as acute myeloid leukemia (AML) and breast cancer (*38*).

Therefore, understanding the precise atomic-level mechanisms driving the formation of these condensates is of therapeutic interest. Studying the water environment in these biological systems will increase our understanding of neurodegenerative disease

development and also allow for the development of targeted therapeutic mechanisms. By unraveling the atomistic details governing the assembly of molecular condensates, novel strategies may emerge with the potential to modulate or disrupt these structures, which might provide a path to disrupt the pathological aggregation of these biomolecules.

Chapter 2

Theoretical background

This theoretical framework draws upon foundational concepts outlined in several texts, including: T. Hill, An Introduction to statistical mechanics, 1986 (*39*); M. Rubinstein, Polymer physics, 2003 (*40*); D. Frenkel, B. Smit, Understanding molecular simulation, 2023 (*41*); F. Menczer, S. Fortunato, C. A. Davies, A first course in network science, 2020 (*42*) and GROMACS documentation section (*43*, *44*).

Biophysics is an interdisciplinary science that applies the principles and techniques of physics to study biological systems. This chapter lays the groundwork for delving into the key principles that underpin the relation between physical concepts and biological systems. The complexity of biological systems requires an interdisciplinary approach, where the rigor of physics is employed to shed light to complex biological processes. Aiming to understand the principles that govern molecular interactions, thermodynamic landscapes, dynamic behavior of biomolecules and provide a theoretical framework for interpreting experimental observations and computational simulations.

2.1 Statistical mechanics

Statistical mechanics is a branch of physics that aims to understand the behavior of macroscopic systems by analyzing the statistical properties of their microscopic constituents. It provides a powerful framework for describing and predicting the thermodynamic properties of gases, liquids, solids, and complex systems like biological systems.

Statistical mechanics is a fundamental tool to understand protein folding pathways and conformational dynamics. By modeling the energy landscapes and probability distributions of protein conformations, it's possible to understand structural transitions of proteins. Statistical mechanics theoretical principles can also be directly applied to create computational models that serve as a tool to understand biological processes, improve experimental design and perform large scale computational analyses.

2.1.1 Microstates and Macrostates

In statistical mechanics, a microstate refers to a specific configuration of the microscopic components like atoms, molecules or particles in a system, characterized by their positions and momenta. A macrostate, on the other hand, describes the overall properties of the system at a macroscopic level, such as temperature, pressure, and volume.

2.1.2 Thermodynamical ensembles and probability distribution

A thermodynamic ensemble is a collection of possible states of a system, each of which has a certain probability of occurring. It provides a theoretical framework for describing the statistical properties of systems in equilibrium. There are several types of thermodynamic ensembles, each used for different types of theoretical systems. In the microcanonical ensemble, the system is isolated and its energy E, volume V, and number of particles N are fixed. The microcanonical ensemble describes systems where the system's energy is conserved.

Its possible to discuss the equivalence of different ensembles through Legendre transformations and the relationship between extensive and intensive quantities.

The microcanonical ensemble is defined by the extensive quantities N, V, and E, which correspond to the number of particles, volume, and total energy of the system, respectively. In contrast, the canonical ensemble is constructed by relaxing the constraint on energy and allowing the system to exchange heat with a reservoir at a fixed temperature T, which is an intensive quantity. The transition from the microcanonical to the canonical ensemble is achieved through a Legendre transformation of the entropy S, converting the internal energy dependence to a dependence on temperature.

$$F(T, V, N) = E - TS, \tag{2.1}$$

where F is the Helmholtz free energy. Here, $T = \left(\frac{\partial E}{\partial S}\right)_{V,N}$ acts as the intensive conjugate variable to the entropy S.

Similarly, the grand canonical ensemble is constructed by allowing the system to exchange particles with a reservoir, fixing the chemical potential μ as the intensive conjugate of the particle number N. This is again achieved through a Legendre transformation, this time involving the Helmholtz free energy:

$$\Omega(T, V, \mu) = F - \mu N = E - TS - \mu N, \tag{2.2}$$

where Ω is the grand potential, and $\mu = \left(\frac{\partial E}{\partial N}\right)_{S,V}$ is the intensive variable associated with N.

The different thermodynamic ensembles (microcanonical, canonical, and grand canonical) are related by Legendre transformations, which enable the interchange of extensive quantities (E, N, V) with their corresponding intensive conjugates (T, μ, P).

The equivalence of ensembles arises in the thermodynamic limit, where fluctuations in the exchanged quantities become negligible relative to the total system size, ensuring that the macroscopic properties predicted by each ensemble converge.

The probability distribution function for the microcanonical ensemble is given by:

$$P(E) = \frac{1}{\Omega(E)} \tag{2.3}$$

where $\Omega(E)$ is the number of microstates corresponding to energy E.

The canonical ensemble describes systems in thermal equilibrium with a heat bath at constant temperature T. In this ensemble, the system is allowed to exchange energy with the surroundings but has fixed volume V and number of particles N.

The central concept in statistical mechanics is the probability distribution function, which gives the probability of finding a system in a particular microstate. In the canonical ensemble, the probability P_i of finding a particle in the i^{th} energy state is given by the Boltzmann factor:

$$P_i = \frac{e^{-\beta E_i}}{Z} \tag{2.4}$$

where:

- P_i is the probability of finding a particle in the i^{th} energy state.
- E_i is the energy of the i^{th} state.
- $\beta = \frac{1}{k_B T}$ is the reciprocal of the product of Boltzmann's constant (k_B) and the temperature (T).

- Z is the partition function, which ensures that the probabilities sum up to unity:

$$Z = \sum_i e^{-\beta E_i} \tag{2.5}$$

The Boltzmann factor $e^{-\beta E_i}$ decreases exponentially with increasing energy, meaning that lower energy states are more likely to be occupied than higher energy states at thermal equilibrium.

The Boltzmann distribution describes the probability of a system occupying a particular energy state. It implies that particles tend to occupy lower energy states more frequently, as these states have higher probabilities according to the Boltzmann factor. It aids in our understanding of the behavior of gases, such as the distribution of molecular speeds in a gas. In solid-state physics, it describes the distribution of electrons among energy bands. I the case of biological systems the Boltzman distribution serves as a link between statistical mechanics and sampled states of a system. Analyzing the probability distribution of molecular states allows us to interpret experimental observations and infer the equilibrium states of biological systems.

The grand canonical ensemble extends the canonical ensemble by allowing both energy and particles to exchange with the surroundings. It describes systems in equilibrium with a reservoir at constant temperature T, chemical potential μ, and volume V. The probability distribution function for the grand canonical ensemble is given by:

$$P = \frac{1}{Z} e^{-\beta(E-\mu N)} \tag{2.6}$$

where Z is the grand canonical partition function.

In principle, any thermodynamic ensemble can be applied to study theoretical systems when considering a large enough system where fluctuations become negligible. But in practice, the choice of ensemble depends on the specific constraints and conditions of the system being studied.

In MD simulations, NVT ensemble is preferred in certain biological contexts where controlling the volume and temperature is more critical than avoiding pressure fluctuations. An example of this, are lipid monolayers membranes, which can be better studied under NVT conditions when their structure is already well-defined, and the goal is to understand their internal dynamics or interactions with surrounding water or solutes. Maintaining a constant volume in NVT simulations helps stabilize the organization of lipids and enables focusing on specific interactions, such as lipid-lipid interactions or how water organizes around lipid heads. Overall, NVT simulations are advantageous in these biological systems because they focus on precise temperature control and internal dynamics, avoiding complications from pressure fluctuations, which may not be biologically relevant or necessary in such contexts.

For systems where a protein is solvated by water or another solvent, NPT simulations are typically employed when is relevant to ensure that the system matches the pressure of an experimental setup. NPT allows for fluctuations in the system's volume to account for solvent compressibility. This is particularly important for studying large, flexible proteins, as the interactions between the protein and the surrounding solvent need to be accurate for understanding folding, dynamics, or solvent-accessible surface areas. Overall the choice of ensemble in MD simulations depends on the desired observables and constraints like temperature, or pressure control, than on intrinsic restrictions of the system itself.

2.1.3 Entropy

Entropy is a fundamental concept in biology, physics, chemistry, and information theory. It is a measure of the disorder or randomness in a system and plays a crucial role in thermodynamics and statistical mechanics. Thermodynamic entropy, statistical mechanics entropy, and information entropy can be understood as different points of view of the same fundamental concept. The measure of disorder, uncertainty, or information in a system. They are equivalent in that they all provide a way to quantify the number of possible microstates or configurations a system can have.

In statistical mechanics this corresponds to the number of microstates that correspond to a particular macroscopic state. In information theory, it's the number of possible outcomes or messages that can be communicated or the uncertainty about which state the system is in.

In biology, however, the concept of entropy is utilized to analyze the diversity of molecular configurations within biological systems, and to quantify molecular disorder. The entropy (S) of a system can be expressed in terms of the partition function as:

$$S = k_B \ln(Z) + k_B \beta \langle E \rangle \tag{2.7}$$

where k_B is Boltzmann's constant, $\langle E \rangle$ is the ensemble average of the system, and $\beta = \frac{1}{k_B T}$ is the reciprocal of the product of Boltzmann's constant and temperature (T). However, other thermodynamic quantities such as internal energy (U), Helmholtz free energy (F), and Gibbs free energy (G) can also be expressed in terms of the partition function.

The entropy of a system can be expressed as:

$$S = k_B \ln(\Omega) \tag{2.8}$$

where k_B is Boltzmann's constant and Ω is the number of microstates corresponding to the macroscopic state of the system. This equation, known as Boltzmann's formula, relates entropy to the multiplicity of possible microscopic arrangements of the system.

The second law of thermodynamics states that the entropy of an isolated system tends to increase over time. This is often expressed as the Clausius inequality:

$$\Delta S \geq \frac{Q}{T} \tag{2.9}$$

where ΔS is the change in entropy, Q is the heat added to the system, and T is the temperature. This inequality implies that in any irreversible process, the total entropy of the system and its surroundings must increase.

In statistical mechanics, entropy is related to the probability distribution of microstates. For a system with discrete energy levels, the entropy can be expressed as:

$$S = -k_B \sum_i P_i \ln(P_i) \tag{2.10}$$

where P_i is the probability of finding the system in the i^{th} microstate.

For a continuous energy spectrum, the entropy is given by:

$$S = -k_B \int P(E) \ln(P(E))\, dE \tag{2.11}$$

Entropy thus quantifies the uncertainty associated with the distribution of microstates in a system.

In information theory, entropy is a measure of the uncertainty or randomness in a set of data. It quantifies the average amount of information produced by a random variable. The entropy of a discrete random variable X with probability mass function $P(X)$ is given by:

$$H(X) = -\sum_{x} P(x) \log_2(P(x)) \tag{2.12}$$

where the sum is taken over all possible values of X. For a continuous random variable, the entropy is defined similarly using the probability density function.

2.2 Solvation

Solvation energy, is the quantity of energy associated with dissolving a solute in a solvent, quantifies the energetic cost or gain associated with solute-solvent interactions.

Solvation is the process by which solvent molecules interact with and surround solute molecules, is a highly relevant phenomenon in biophysics that influences the stability, structure, and dynamics of biomolecular systems. The interaction between solvent molecules and biomolecular surfaces influences the stability and reactivity of macromolecules such as proteins, and lipids. Understanding solvation energies is crucial for deciphering the molecular forces that govern biomolecular interactions in physiological solution, and play a vital role in multiple biological processes, such as protein folding and protein agregation. The hydrophobic effect, driven by the tendency of non polar molecules to minimize contact with water, is a prominent contributor for solvation energies in proteins. The release of water molecules from hydrophobic aminoacids during protein folding contributes to a favorable solvation energy, influencing the overall stability of the folded protein structure.

2.2.1 Solvation Energy

Solvation energy, denoted by ΔG_{solv}, is the change in Gibbs free energy associated with the solvation process. It can be expressed as:

$$\Delta G_{\text{solv}} = \Delta H_{\text{solv}} - T\Delta S_{\text{solv}} \tag{2.13}$$

where:

- ΔH_{solv} is the enthalpy change of solvation, representing the energy released or absorbed when solute molecules interact with solvent molecules.
- ΔS_{solv} is the entropy change of solvation, reflecting the increase or decrease in randomness associated with the solvation process.
- T is the temperature.

2.2.2 Factors Influencing Solvation Energy

The solvation energy of a solute depends on various factors, including:

1. **Nature of solvent:** Different solvents have different solvation energies due to variations in polarity, hydrogen bonding capability, and dielectric constant.
2. **Nature of solute:** The size, charge, and polarity of the solute molecule influence its interaction with solvent molecules and hence the solvation energy.
3. **Temperature:** Solvation energy is temperature-dependent, with both enthalpy and entropy contributions varying with temperature.
4. **Concentration:** Solvation energy may exhibit concentration dependence, especially in cases of solute-solute interactions or solvent-solvent interactions.
5. **Pressure:** High pressures can affect solvation energy by altering the solvent density and compressibility.

2.2.3 Phase separation theory

Phase separation theory describes the spontaneous division of a homogeneous mixture into two or more distinct phases. This phenomenon is observed in various systems, including liquids, polymers, and colloidal suspensions. Phase separation occurs due to minimization of the free energy of the system (ΔG). It occurs when the free energy change is negative ($\Delta G < 0$), indicating that the system can achieve a lower energy state by forming distinct phases.

The specific mechanism of phase separation depends on factors such as temperature, pressure, composition, and interactions between molecules (*45*). Phase separation theory provides a framework for understanding the spontaneous division of homogeneous mixtures into distinct phases. This fundamental concept in physics has gained attention in the field of biophysics due to its role in organizing cellular compartments and influencing cellular functions (*45*). This phenomenon permits the creation of membraneless organelles or biomolecular condensates. These membraneless compartments play a vital role in cellular processes, by bringing localized concentrations of biomolecules and facilitating specific biochemical reactions.

In biological systems phase separation is driven by the balance of attractive and repulsive forces among biomolecules. Weak multivalent interactions, such as protein-protein interactions in IDPs contribute to the formation of liquid-liquid phase-separated condensates. The dynamic nature of these interactions allows for the rapid and reversible assembly and disassembly of condensates.

There are several examples of phase separated biomolecular condensates, each with distinct functions. Stress granules and P-bodies are involved in cellular stress responses and mRNA regulation. Nuclear bodies, such as nucleoli, contribute to ribosomal RNA processing.

Phase-separated condensates have been implicated in diverse cellular functions, including gene expression regulation, signal transduction, and the response to cellular stress. Dysregulation of phase separated biomolecular condensates are associated with various diseases, including neurodegenerative disorders and certain cancers. Targeting phase-separated condensates has gained importance due to its importance for drug development, particularly in diseases where aberrant phase separation contributes to disease.

2.2.4 Phase separation of polymers

Flory-Huggins

The Flory-Huggins theory is a fundamental model in polymer physics that describes the thermodynamics of polymer solutions and polymer blends. Developed independently by Paul Flory and Michael Huggins in the 1950s (*46*, *47*), this theory provides insights into the mixing behavior of polymers and predicts phase separation phenomena.

To derive the Flory-Huggins model is necessary to take into account the following considerations:

- The solution is modeled on a lattice, with each lattice site occupied either by a solvent molecule or a polymer segment.
- The polymer chains are made up of several segments, each occupying one lattice site.
- The mixing of polymer segments and solvent molecules occurs randomly, and interactions are localized between adjacent lattice sites.
- The polymer-solvent interactions are described by the Flory-Huggins interaction parameter, χ, which accounts for the energetic interactions between polymer and solvent.

The derivation of the Flory-Huggins model consists in considering the entropic and enthalpic contributions to the free energy of mixing. The free energy of mixing is derived from the lattice model of mixing, first we consider the entropy of mixing.

The entropy of mixing ΔS_m for a system consisting of a polymer and a solvent, arises from the number of possible configurations of polymer segments and solvent molecules on a lattice.

Where:

- N_1 be the number of solvent molecules, with each molecule occupying one lattice site.
- N_2 be the number of polymer molecules, with each polymer composed of r segments (so the total number of polymer segments is rN_2).
- ϕ_1 and ϕ_2 be the volume fractions of the solvent and polymer, respectively.

The entropy of mixing is given by the number of distinct ways to arrange solvent molecules and polymer segments on a lattice:

$$\Delta S_m = -k_B \left(N_1 \ln \phi_1 + \frac{N_2}{r} \ln \phi_2 \right) \tag{2.14}$$

Where:

- k_B is the Boltzmann constant,
- $\phi_1 = \frac{N_1}{N_1 + rN_2}$ is the volume fraction of solvent,
- $\phi_2 = \frac{rN_2}{N_1 + rN_2}$ is the volume fraction of polymer.

This equation accounts for the configurational entropy lost when polymer chains, which are much larger than solvent molecules, are arranged within the solvent. Due to the connectivity of polymer chains, the entropy of mixing is significantly reduced compared to a system of small molecules.

Next, we consider the enthalpy of mixing ΔH_m, which is determined by the interactions between polymer segments and solvent molecules. The Flory-Huggins interaction parameter χ describes the energy difference between polymer-solvent and solvent-solvent or polymer-polymer interactions.

The enthalpy of mixing is given by:

$$\Delta H_m = \chi_{12} \phi_2 N_1 k_B T \tag{2.15}$$

Where:

- χ is the Flory-Huggins interaction parameter,

- N is the total number of lattice sites,
- T is the temperature.

The interaction parameter χ can be positive for unfavorable polymer-solvent interactions or negative for favorable polymer-solvent interactions. Because of this a positive χ promotes phase separation, while a negative χ favors mixing.

The total free energy of mixing ΔG_m is the sum of the entropic and enthalpic contributions:

$$\Delta G_m = \Delta H_m - T\Delta S_m \tag{2.16}$$

Substituting the expressions for ΔS_m and ΔH_m, we obtain the Flory-Huggins free energy of mixing per lattice site:

$$\Delta G_m = k_B T \left(\frac{\phi_1}{r} \ln \phi_1 + \phi_2 \ln \phi_2 + \chi \phi_1 \phi_2 \right) \tag{2.17}$$

Where:

- ϕ_1 and ϕ_2 are the volume fractions of the solvent and polymer, respectively,
- r is the number of segments per polymer chain,
- χ is the Flory-Huggins interaction parameter.

To move to a macroscopic (per mole) description, we scale the free energy by the number of moles of molecules or segments.

Where:

- N_1 be the number of solvent molecules,
- N_2 be the number of polymer molecules,
- r be the number of segments per polymer molecule.

Thus, the total number of lattice sites N is:

$$N = N_1 + rN_2 \tag{2.18}$$

This means that the number of lattice sites occupied by the polymer is proportional to the number of segments in each polymer molecule, $(r \times N_2)$.

To convert ΔG_m from per segment to per mole, we multiply by the total number of lattice sites:

$$\Delta G_{\text{mix}} = N_A \cdot \Delta G_m \tag{2.19}$$

Substituting for ΔG_m:

$$\Delta G_{\text{mix}} = N_A k_B T \left(\frac{\phi_1}{r} \ln \phi_1 + \phi_2 \ln \phi_2 + \chi \phi_1 \phi_2 \right) \tag{2.20}$$

We then use the fact that $N_A k_B = R$, where R is the gas constant. Thus, the equation becomes:

$$\Delta G_{\text{mix}} = RT \left(\frac{\phi_1}{r} \ln \phi_1 + \phi_2 \ln \phi_2 + \chi \phi_1 \phi_2 \right) \tag{2.21}$$

Now, we express the volume fractions ϕ_1 and ϕ_2 in terms of mole numbers:

$$\phi_1 = \frac{n_1}{n_1 + rn_2}, \quad \phi_2 = \frac{rn_2}{n_1 + rn_2} \tag{2.22}$$

Finally, the macroscopic equation for the free energy of mixing becomes:

$$\Delta G_{\text{mix}} = RT \left(n_1 \ln \phi_1 + n_2 \ln \phi_2 + n_1 \chi \phi_1 \phi_2 \right) \tag{2.23}$$

where n_1 is the number of moles of solvent and n_2 is the number of moles of polymer.

By analyzing the Flory-Huggins free energy expression, one can predict the phase behavior of polymer blends at different temperatures and compositions. For instance, when χ exceeds a critical value, phase separation occurs, leading to the formation of two distinct phases, as shown in Figure 2.1. The phase diagram can be constructed

based on the Flory-Huggins theory, providing valuable insights into the miscibility of polymer blends.

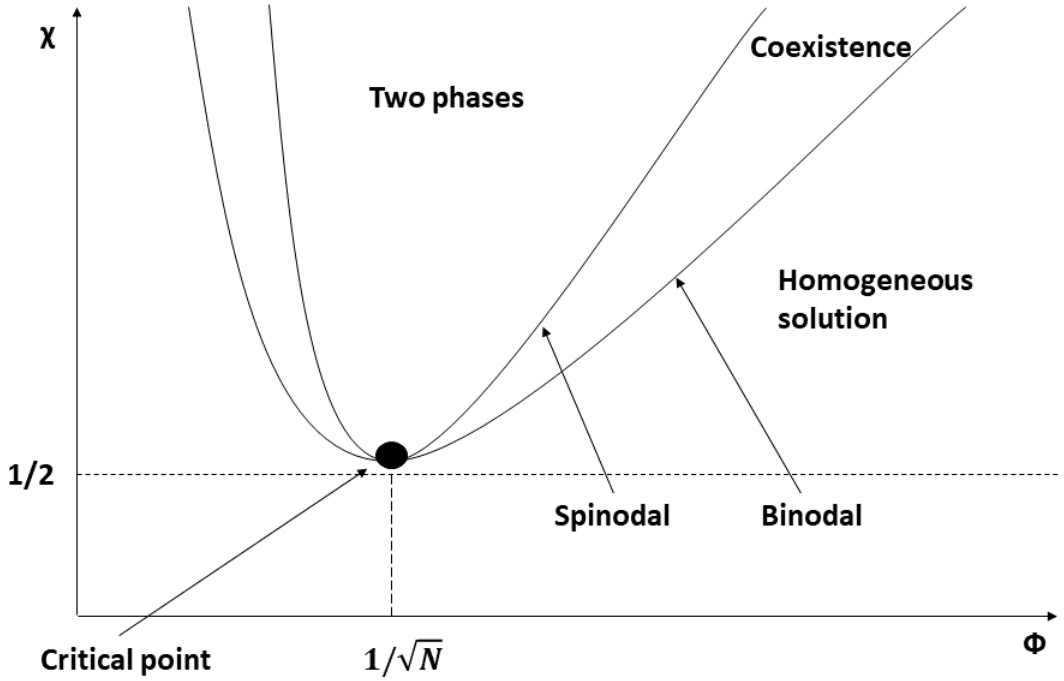

Figure 2.1: **Phase diagram** of phase separation displaying the critical point with the filled black point, where phase separation does not occur. Depending of the values of ϕ and χ (X and Y axis) the system will be in homogeneous, coexistence or phase separated state. The diagram illustrates the binodal and spinodal curve. The region between the binodal and spinodal curves corresponds to the coexistence phase. In the region enclosed by the spinodal curve, phase separation will occur.

Phase separation in polymer mixtures can occur through different mechanisms depending on the thermodynamic conditions. One of the mechanisms is spinodal decomposition, which occurs when the mixture is thermodynamically unstable with respect to small perturbations in composition. In spinodal decomposition, the system undergoes spontaneous phase separation without the formation of a distinct interface between the phases. Another mechanism is binodal decomposition, which occurs when the mixture is thermodynamically metastable with respect to small perturbations in composition. In binodal decomposition, phase separation occurs through the nucleation

and growth of small domains of one phase within the matrix of the other phase, leading to the formation of a distinct interface between the phases.

Phase separation in polymer mixtures is governed by various factors such as polymer-polymer interactions, temperature, and composition. In polymer solutions, liquid-liquid phase separation occurs when the concentration of polymers surpasses a critical point, resulting in the formation of coexisting polymer-rich and polymer-poor phases. This phenomenon occurs in biological systems, one example of this phenomena lies within intracellular environments, where macromolecules like proteins and nucleic acids can liquid-liquid phase separate forming membraneless organelles.

2.2.5 Polymer theory for biological systems

Polymer physics provides an ideal theoretical framework for understanding biological processes, offering insights into the behavior of biopolymers and synthetic polymers within living systems. Polymer theory concepts, such as the Flory–Huggins theory, allows to analyze the equilibrium conformations, structural transitions, and stability of biopolymers.

In the realm of biological processes, concepts like Gibbs free energy, enthalpy and entropy calculations based on polymer theory allow examining the energy of binding events, such as protein–protein interactions or DNA–protein recognition. Polymer physics models, and chain models, are used to understand the behavior of biopolymers like actin and microtubules or IDP proteins.

The phenomenon of phase separation in biology, typically associated with cellular organization and biomolecular condensation, has gained increasing attention for its potential implications in the onset and progression of various diseases, aberrant cellular functions and the formation of pathological structures.

Understanding the involvement of phase separation and nucleation processes in disease pathogenesis is relevant for developing new therapeutic approaches. Targeting specific phases in these processes, such as disrupting protein aggregation or modulating cellular phase separation, could open the way for developing innovative treatments for diseases where these processes are involved.

2.3 Computational methods

2.3.1 Molecular dynamics simulations

Molecular Dynamics (MD) simulations are a powerful tool in biophysics, allowing for the exploration of the dynamic behavior of biomolecules at an atomistic resolution. MD simulations numerically solve the equations of motion for a system of interacting atoms or molecules over time. With these, basic principles from classical mechanics, statistical mechanics, and quantum mechanics can be used depending on the scale of the simulated system. Relying on force fields to describe interatomic interactions and the integration algorithms used to propagate the system's dynamics. Statistical mechanics provides the theoretical framework for connecting microscopic properties of a system with its macroscopic behavior. In MD simulations, statistical mechanics principles are used to calculate ensemble averages of various thermodynamic quantities.

In MD simulations, the ensemble refers to the collection of all possible states that the system can occupy under specified conditions of temperature, volume, and number of particles. In this dissertation the ensembles used for the MD simulations are the canonical ensemble (NVT), and isothermal-isobaric ensemble (NPT).

In biology, MD simulations can be used to understand the complexities of protein folding and dynamics. By simulating the trajectories of atoms over nanosecond to microsecond timescales, it's possible to gain insights into the folding pathways, stability, and conformational changes of proteins. Another example of of MD simulations is to

provide a microscopic view of DNA and RNA dynamics, understanding the structural dynamics of DNA during processes like replication, transcription, and repair.

MD simulations also play a role in elucidating the dynamics of lipid bilayers and their interactions with membrane proteins. By simulating systems of lipids in membranes, the formation of lipid rafts and the dynamics of membrane-bound proteins has been used to explore cellular membrane functions. MD simulations also enable the study of the solvation and hydration of biomolecules which is crucial for their stability and functionality. MD simulations of water molecules around biomolecules, enhances understanding of solvation energies, hydration shells, and the effects of water on protein folding.

2.3.2 Thermodynamic Variables

Force fields

In our simulations we employ a classical force field approximation, where interactions between atoms are described using empirical mathematical functions. These functions are based on principles of classical mechanics and are parameterized to capture the behavior of molecular systems. Importantly, this classical approximation does not account for quantum mechanical effects, and instead relies on approximations derived from experimental data and simplified classical physics models for atomic interactions.

I these force fields the total energy E of a molecular system in an MD simulation is typically calculated as the sum of different energy contributions:

1. **Bond energy ($E_{\mathbf{bond}}$)**: This term accounts for the energy associated with the bonds between atoms in the system. It includes contributions from covalent bonds, angle bending, and dihedral rotations. The bond energy is often described using harmonic potentials for simple bonds.

For a system with N_{bonds}, the bond energy can be calculated as:

$$E_{\text{bond}} = \sum_{i=1}^{N_{\text{bonds}}} k_i (r_i - r_{i,0})^2 \tag{2.24}$$

where k_i is the force constant of the i-th bond, r_i is its current length, and $r_{i,0}$ is its equilibrium length.

2. **Non-bonded energy ($E_{\text{non-bonded}}$)**: Non-bonded interactions refer to interactions between atoms that are not directly bonded to each other. These interactions include van der Waals forces (dispersion forces) and electrostatic interactions (Coulombic forces). The van der Waals interactions are often modeled using Lennard-Jones potentials, while electrostatic interactions are described by the Coulomb potential.

   The non-bonded energy can be represented as:

$$E_{\text{non-bonded}} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \tag{2.25}$$

   where σ_{ij} is the parameter of the Lennard-Jones potential, q_i and q_j are the charges of atoms i and j, ε is the dielectric constant, and r_{ij} is the distance between atoms i and j.

3. **External potential energy (E_{external})**: If the system is subjected to external fields or potentials (such as a confinement potential, the energy associated with these external influences is also included in the total energy calculation.

   This can be represented as:

$$E_{\text{external}} = \sum_{i=1}^{N} U_{\text{ext}}(r_i) \tag{2.26}$$

   where $U_{\text{ext}}(r_i)$ is the external potential acting on atom i.

4. **Kinetic energy (K)**: The kinetic energy of the system arises from the motion of atoms and molecules. In MD simulations, it is typically calculated using the velocities of particles according to classical mechanics, where kinetic energy K is given by $K = \frac{1}{2}mv^2$, where m is the mass of the particle and v is its velocity.

The total energy of the system, E_{total}, is then the sum of all these contributions:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{non-bonded}} + E_{\text{external}} + K \tag{2.27}$$

In MD simulations, this total energy is typically calculated at each time step to monitor the system's behavior, ensure energy conservation and analyze thermodynamic properties.

Temperature

In molecular dynamics simulations, temperature is calculated by the average kinetic energy of the particles in the system. It is calculated using the velocities of the particles according to the equipartition theorem.

The temperature T of the system can be calculated using the average kinetic energy per degree of freedom:

$$T = \langle \frac{2}{3k_B} K \rangle \tag{2.28}$$

where:

- T is the temperature,
- k_B is the Boltzmann constant ($k_B \approx 1.38 \times 10^{-23}\,\text{J/K}$),
- K is the total kinetic energy of the system.

In practice, the istantaneous temperature is calculated periodically during the simulation by averaging the kinetic energies of all particles and using equation 2.28. MD thermostats are used in order to reach the desired temperature of the system an example of these thermostats is the Bussi velocity rescaling algorithm (*48*) used in

the simulations performed for this dissertation. The Bussi velocity rescaling algorithm operates as follows:

The thermostat adjusts the velocities based on the difference between the instantaneous temperature T and the target temperature T_{target}. The equation for calculating the deviation of the system temperature to the target temperature is:

$$\frac{d\mathbf{T}}{dt} = -\frac{1}{\tau}(\mathbf{T}_0 - \mathbf{T}) \tag{2.29}$$

where:

- $\mathbf{T}_0$ is the target temperature of the system,
- $\mathbf{T}$ is the instantaneous temperature,
- τ is the temperature coupling time constant, which determines how quickly the system equilibrates to the target temperature.

In this thermostat, the equation is implemented using a velocity rescaling approach where the velocity of each particle is scaled according to a factor:

$$\lambda = \sqrt{1 + \frac{\Delta t n_{TC}}{\tau}\left(\frac{T_0}{T\left(t - \frac{\Delta t}{2}\right)} - 1\right)} \tag{2.30}$$

where:

- Δt is the time step used in the simulation,
- T is the instantaneous temperature of the system,
- n_{tc} refers to each step,
- and τ the time constant

Then the thermostat modifies the kinetic energy at each scaling step by:

$$\mathbf{\Delta}E_k = (\lambda - 1)^2 E_k \tag{2.31}$$

The sum of these changes over the simulation run needs to be subtracted from the total energy to obtain the conserved energy quantity. Nevertheless this scaling is not thermodynamically correct, as it does not preserve the ensemble in the long term, but it is often used for temperature equilibration in initial simulation stages. To produce the correct ensemble we add an stochastic term that ensures a correct kinetic energy distribution by modifying it according to:

$$\mathbf{d}K = (K_0 - K)\frac{dt}{\tau} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau}} \tag{2.32}$$

where K is the kinetic energy, N_f the number of degrees of freedom and dW a Wiener process.

The velocity rescaling algorithm is then applied periodically during the simulation to maintain temperature control. By adjusting the velocities of particles based on the difference between the current and target temperatures, this algorithm ensures that the simulated system matches the desired thermodynamic conditions.

Pressure

Pressure (P) represents the force per unit area exerted by the particles on the boundaries of the system. In MD simulations, pressure control is essential for simulations performed under constant pressure conditions (NPT ensemble). Barostats are used to maintain the desired pressure by adjusting the volume of the simulation cell.

Barostat algorithms are used to control the pressure of the simulated system. These algorithms mimic the role of a pressure bath in a real-world scenario, ensuring that the system maintains a desired pressure throughout the simulation. Barostats are particularly important when studying systems under constant pressure conditions or when simulating phase transitions. Some barostat algorithms are:

Berendsen Barostat

The Berendsen barostat is a simple and widely used method for pressure control in MD simulations. It operates by scaling the box dimensions (volume) of the simulation cell based on the difference between the current pressure and the target pressure. This scaling factor is gradually applied over a relaxation time to achieve the desired pressure.

1. **Calculate current pressure**: First, the instantaneous pressure P_{current} of the system is calculated. This is typically done using the virial stress tensor or other pressure estimators based on the positions and velocities of particles.

2. **Calculate scaling factor**: Next, a scaling factor s is determined based on the difference between the current pressure P_{current} and the target pressure P_{target}. This scaling factor adjusts the volume of the simulation cell to rescale the pressure towards the desired value. The scaling factor s is often calculated as:

$$\mu = 1 + \frac{k_T \Delta t}{3\tau}(P_{\text{target}} - P_{\text{current}}) \tag{2.33}$$

   where Δt is the simulation time step, k_T is the compressibility, and τ is a relaxation time constant that controls the rate of volume scaling.

3. **Rescale volume**: Finally, the volume of the simulation cell is scaled by the factor μ to adjust the pressure towards the target pressure P_{target}. This rescaling process ensures that the system maintains the desired pressure throughout the simulation.

The Berendsen barostat algorithm is effective for maintaining pressure control,the Berendsen barostat does not yield the exact NPT ensemble. Instead, it provides a simple and efficient method for quickly equilibrating the system to the desired pressure conditions.

Parrinello-Rahman Barostat:

The Parrinello-Rahman barostat is a more sophisticated method that correctly yields the NPT ensemble. It introduces additional degrees of freedom to the system, allowing the box dimensions to evolve dynamically in response to changes in pressure. This barostat is particularly useful for simulating systems undergoing large volume fluctuations, such as phase transitions or structural transformations.

The Parrinello-Rahman method modifies the simulation box vectors **h**, which describe the size and shape of the simulation box. The time evolution of **h** is determined by the equation:

$$\frac{d\mathbf{b^2}}{dt^2} = VW^{-1}b^{`-1}\left(\mathbf{P} - \mathbf{P}_{\text{ref}}\right) \tag{2.34}$$

The volume of the box is denoted V, and W is a matrix parameter that determines the strength of the coupling. The matrices and $_{r}ef$ are the current and reference pressures, respectively.

The modified Hamiltonian will be:

$$\mathbf{E_{pot}} + \mathbf{E_{kin}} = \sum_{i} P_{ii}\mathbf{V} + \sum_{i,j} \frac{1}{2}\mathbf{W}_{ij}(\frac{db_{ij}}{dt})^2 \tag{2.35}$$

The equations of motion for the atoms, obtained from the Hamiltonian are:

$$\frac{\mathbf{d^2r_i}}{\mathbf{dt^2}} = \frac{\mathbf{F_i}}{\mathbf{m_i}} - \mathbf{M}\frac{dr_i}{dt} \tag{2.36}$$

$$\mathbf{M} = \mathbf{b^{-1}}\left[b\frac{d^{`}}{dt} + \frac{\mathbf{db}}{\mathbf{dt}}\mathbf{b}^{`}\right]b^{`-1} \tag{2.37}$$

This extra term might look like friction, but it's actually a result of how the Parrinello-Rahman equations are defined. These equations use particle coordinates relative to the box vectors, while most MD simulations use standard Cartesian coordinates for positions, velocities, and forces.

The Parrinello-Rahman barostat algorithm provides an efficient and accurate method for pressure control in MD simulations, particularly for systems undergoing large volume fluctuations or structural transformations. By conserving both energy and volume fluctuations, it ensures that the simulated system closely matches the desired thermodynamic conditions.

Barostats play a critical role in ensuring that MD simulations accurately capture the behavior of real-world systems at specified pressure conditions. Each barostat algorithm has its advantages and limitations, depending on the specific characteristics of the system being simulated and the desired balance between accuracy, efficiency, and conservation of thermodynamic properties.

2.3.3 Dynamics algorithms

Velocity verlet algorithm

The velocity Verlet algorithm is a numerical integration scheme used in simulations for solving the equations of motion. It is an extension of the original Verlet algorithm that includes velocities explicitly and is used for its simplicity, efficiency, and conservation of energy.

The velocity Verlet algorithm computes the positions and velocities of particles at each time step based on their accelerations and initial conditions. It works as follows:

1. **Initialization**: At the beginning of the simulation, initial positions $\mathbf{r}_i(t_0)$ and velocities $\mathbf{v}_i(t_0)$ of all particles are known.

2. **Position update**: The positions of particles are advanced by a full time step (Δt) using the current velocities and accelerations:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i(t) + \frac{(\Delta t)^2}{2m} \cdot \mathbf{F}_i(t) \tag{2.38}$$

   where $\mathbf{F}_i(t)$ is the force acting on particle i at time t, m is the mass of the particle, and Δt is the time step.

3. **Velocity half-step**: The velocities are updated by a half time step ($\Delta t/2$) using the current accelerations and the average of accelerations at the current and updated positions:

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t) + \frac{\Delta t}{2m} \left(\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)\right) \tag{2.39}$$

4. **Force recalculation**: Using the updated positions, forces are recalculated based on the new configuration of particles.

5. **Velocity update**: Finally, the velocities are updated to the full time step using the average of the accelerations at the current and updated positions:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t + \Delta t/2) + \frac{\Delta t}{2m} \left(\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)\right) \tag{2.40}$$

Leapfrog Algorithm

The leapfrog algorithm is another numerical integration scheme used in molecular dynamics simulations for solving the equations of motion. This algorithm computes the positions and velocities of particles at each time step based on their accelerations and initial conditions in the following way:

1. **Initialization**: At the beginning of the simulation, initial positions $\mathbf{r}_i(t_0)$ and velocities $\mathbf{v}_i(t_0)$ of all particles are known.

2. **Half-step velocity update**: The velocities of particles are updated by a half time step ($\Delta t/2$) using the current accelerations:

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t - \frac{\Delta t}{2}) + \frac{\Delta t}{m} \cdot \mathbf{F}_i(t) \tag{2.41}$$

where $\mathbf{F}_i(t)$ is the force acting on particle i at time t, m is the mass of the particle, and Δt is the time step.

3. **Position update**: The positions of particles are advanced by a full time step (Δt) using the updated velocities:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i(t + \Delta t/2) \tag{2.42}$$

4. **Force recalculation**: Using the updated positions, forces are recalculated based on the new configuration of particles.

5. **Full-step velocity update**: Finally, the velocities are updated to the full time step using the new accelerations:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t + \Delta t/2) + \frac{\Delta t}{2m} \cdot \mathbf{F}_i(t + \Delta t) \tag{2.43}$$

2.3.4 Coarse-grained models and backmapping schemes

Coarse-grained (CG) models and backmapping schemes are techniques used in molecular simulations in order to bridge the gap between computational efficiency and molecular detail.

Coarse graining

Coarse-grained models allow simulating larger timescales by reducing the system's degrees of freedom, grouping atoms into simplified "beads," and smoothing energy landscapes. This enables larger time steps and faster computations by averaging out fine-scale fluctuations while focusing on essential large-scale dynamics, facilitating the exploration of biologically relevant phenomena. To properly take advantage of coarse-grained models it is necessary to select the pertinent level of detail, preserve

relevant key interactions, and capture properties of interest for the system of study while minimizing computational costs.

Various coarse-grained models exist, for example united-atom, bead-spring and multiscale models. United-atom models merge neighboring atoms into a single interaction site, reducing the number of degrees of freedom. Bead-spring models represent polymers or macromolecules as connected beads, emphasizing chain connectivity. Multiscale models combine different levels of resolution within a single simulation, providing a versatile approach to study systems focusing on the relevant part of the system while preserving crucial interactions with its environment.

For biological systems, coarse grained models are used when the large size and timescales often exceed the capabilities of all-atom simulations. From simulating membrane dynamics to investigating protein folding pathways and understanding nucleic acid behavior, coarse-graining is a powerful tool to explore biological phenomena that would be computationally prohibitive with finer resolution models.

Backmapping

Backmapping is the process of reverting a coarse-grained representation back to atomistic detail, a crucial step for analyzing processes at the atomistic level. However, backmapping is inherently challenging because coarse-graining is not an injective transformation; that is, multiple atomistic configurations can correspond to the same coarse-grained structure. Additionally, the coarse-grained representation is subjective, as the choice of mapping scheme and level of coarse-graining affect the information retained. Various approaches have been developed to tackle this, including statistical methods that rely on probability distributions, optimization algorithms that minimize potential energy or other metrics, and machine learning techniques trained to predict atomistic structures from coarse-grained representations.

2.4 Graph analysis formalism

Graph theory is a branch of mathematics that, in the context of biology, allows the representation of biological systems as graphs composed of nodes and edges. In our work nodes represent biomolecules (e.g., proteins, amino acids or water molecules), and edges denote the interactions between them, its possible to assign weights to this edges to represent the strength of their interactions. The application of graph theory allows for the representation of biological networks, providing quantitative means to analyze connectivity patterns.

A graph G = (V, e) is defined by a set of nodes V (or vertices) and edges e (see Figure 2.2), where an edge represents a connection between two nodes. In our work, this formalism is used to model a molecular network by representing individual molecules (or specific groups, such as amino acids or water molecules) as nodes, and their interactions as edges.

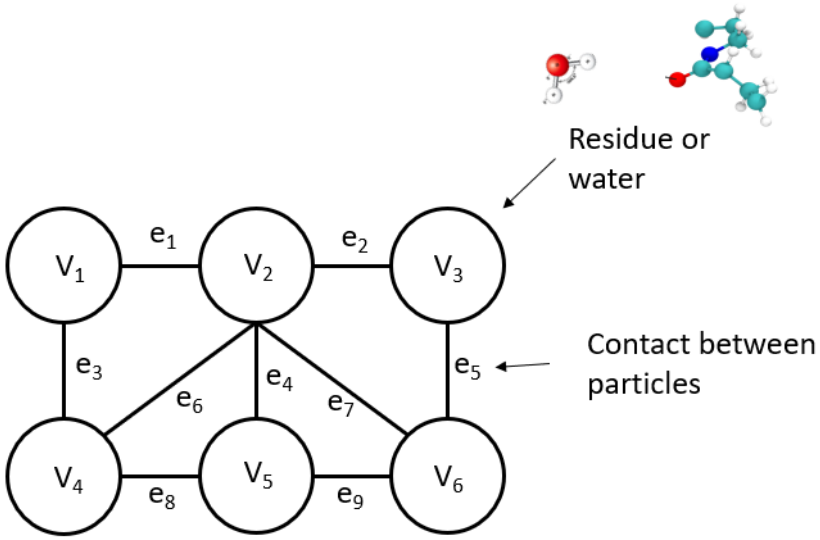

Figure 2.2: **Graph representation** of residues or water molecules as graph nodes V_i and contacts between them as edges e_i

.

The edges in graphs can be weighted and directed (or undirected) depending on the type of interactions being modeled:

- Weighted edges: The weights represent the strength of interaction, such as bond proximity.

- Directed edges: In cases where the interaction has a directional component, such as donor-acceptor relationships in hydrogen bonding, the edges are directed.

- Undirected edges: For symmetrical interactions, such as proximity-based contacts, edges are undirected.

In the network created for the work in this dissertation we took the following considerations.

1. Nodes correspond to entities of interest (e.g., water molecules or amino acids). Each node is assigned attributes such as its identity, protein chain, or degree of interaction.

2. Edges are established based on specific criteria. For instance, edges may be drawn if the distance between two nodes is below a cutoff radius (e.g., 3.5 Å for molecular interactions). Edge weights may be assigned based on metrics such as bond distance.

3. The graph is mathematically represented by an adjacency matrix A, where A_{ij} denotes the presence (and weight, if applicable) of an edge between nodes i and j.

Additionally, key graph-theoretical measures are used to analyze the network structure:

- Degree Centrality:

The degree centrality d_i of a node i in an undirected graph is defined as the number of edges connected to that node:

$$d_i = \sum_{j \in V} A_{ij}, \tag{2.44}$$

where A_{ij} is the element of the adjacency matrix A representing the edge between nodes i and j. For directed graphs, degree centrality can be divided into:

- In-degree: The number of edges directed toward the node:

$$d_i^{\text{in}} = \sum_{j \in V} A_{ji}. \tag{2.45}$$

- Out-degree: The number of edges originating from the node:

$$d_i^{\text{out}} = \sum_{j \in V} A_{ij}. \tag{2.46}$$

Degree centrality provides insight into the connectivity of a node, identifying highly connected nodes that may play critical roles in the network.

- Clustering Coefficient. The clustering coefficient C_i quantifies the tendency of a node i to form tightly connected local neighborhoods. It is defined as the ratio of the number of edges between the neighbors of i to the maximum number of such edges:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}, \tag{2.47}$$

where: - k_i is the degree of node i, - e_i is the number of edges between the neighbors of i.

For a graph as a whole, the average clustering coefficient is calculated as:

$$C = \frac{1}{|V|} \sum_{i \in V} C_i, \tag{2.48}$$

where $|V|$ is the total number of nodes in the graph. The clustering coefficient highlights the presence of local clusters or communities within the network.

- Path Length and Connectivity.

The shortest path length l_{ij} between two nodes i and j is the minimum number of edges that must be traversed to connect i to j:

$$l_{ij} = \min\{\text{number of edges between } i \text{ and } j\}. \tag{2.49}$$

The average path length L for the entire graph is given by:

$$L = \frac{1}{|V|(|V|-1)} \sum_{i \neq j \in V} l_{ij}, \tag{2.50}$$

where the summation is over all pairs of nodes in the graph. The average path length provides a measure of the overall connectivity and compactness of the graph.

Graph analysis formalism can be used to perform topological analysis in order to reveal key features of biological networks. From this topological analysis it is possible to identify crucial nodes and decipher hierarchical structures within biomolecular interactions (*49*).

Graph analysis formalism can be extended to study dynamic changes and evolution in biological networks. Temporal aspects, such as changes in network topology over time and identification of network motifs, provides insight into the dynamic nature

of biological processes. This formalism is employed to understand folding patterns of biomolecules and the interaction interfaces within macromolecular complexes. Through the identification of relevant nodes in biological networks it is possible to target and prioritize molecules of interest to aid in the development of therapeutic drugs.

By combining graph theory with atomistic-level simulations, we are able to analyze the large amounts of data obtained from our simulations in a structured fashion and understand structural elements and organization of large scale simulations.

2.5 Spectroscopy techniques

MD simulation methods provide insights that enhance relevance of experimental results. While MD simulations offer detailed, atomistic views of molecular interactions and dynamic processes, they rely on theoretical models. Spectroscopy techniques such as two dimensional infrared spectroscopy and sum frequency generation spectroscopy allow for direct experimental observation of conformational changes and time dependent behaviors.

Spectroscopy and molecular dynamics simulations are highly complementary techniques. When used together, they provide a clear picture of molecular behavior. Spectroscopy offers experimental validation and data on longer time scales, while MD provides atomistic detail and predictions of molecular motions, enabling understanding of both the structure and dynamics of biomolecules.

2.5.1 Two dimensional infrared spectroscopy

Two-dimensional infrared (2D-IR) spectroscopy offers a unique perspective on molecular dynamics and interactions, giving insights into how molecules move and interact on very short timescales. Unlike traditional infrared (IR) spectroscopy, which measures the absorption of infrared light at different frequencies, 2D-IR spectroscopy adds a second dimension by introducing a sequence of ultrashort infrared light pulses.

These pulses interact with the sample in a timed order, creating a signal that reflects how vibrational modes (the ways molecules vibrate) are coupled or related to each other.

The resulting 2D spectrum is like a map, showing the frequencies of the vibrations along one axis and how they change or interact over time along the other.

Interpreting 2D-IR spectra involves assigning peaks to specific vibrational modes and understanding their temporal evolution. Quantum mechanical calculations and modeling aid in assigning spectral features to specific molecular vibrations, which are then used to gain information of the structural dynamics of biomolecules. The cross-peaks in the 2D spectra reveal vibrational couplings and correlations. 2D-IR spectroscopy is used in structural biology to probe the dynamics of biomolecules. It has been employed to study protein folding, elucidate the dynamics of peptides and proteins in membranes, and investigate the structure and behavior of nucleic acids (*50*). The technique provides a time-resolved view of conformational changes, hydrogen bonding dynamics, and solvent interactions.

Of notable interest for this dissertation is the application of 2D-IR spectroscopy to studying solvation dynamics and water interactions with biomolecules. This technique allows the exploration of the dynamics of water molecules surrounding proteins, providing insights into hydration shells, water-mediated structural changes, and the influence of solvent on biomolecular dynamics.

Time resulution, inherent to 2D-IR spectroscopy, enables the study of ultrafast processes and kinetics in biomolecular systems. Time-resolved experiments have been employed to investigate processes such as energy transfer in photosynthetic systems, ligand binding kinetics, and the dynamics of enzymatic reactions, providing a detailed temporal understanding of these events.

2.5.2 Sum frequency generation spectroscopy

Sum Frequency Generation (SFG) spectroscopy probes molecular organization and interactions at interfaces by exploiting the nonlinear optical properties of materials. It is a second-order nonlinear optical technique, where two input beams, infrared (to excite molecular vibrations) and visible (to induce electronic resonance), interact within a material, generating a signal at the sum of their frequencies. The strength of the SFG signal depends on the second-order nonlinear susceptibility ($\chi^{(2)}$), which is inherently sensitive to the symmetry properties of the system. Importantly, $\chi^{(2)}$ is nonzero only in regions lacking inversion symmetry, such as interfaces, making SFG uniquely suited for studying surface specific phenomena. By analyzing the SFG spectrum, it's possible to connect molecular vibrations and orientations to material properties like interfacial structure, bonding, and surface composition. With this technique, vibrational modes at interfaces, such as CH, NH, and OH stretching vibrations, can be selectively probed, offering insights into molecular structures and orientations.

SFG spectroscopy can be used to study biological interfaces, like lipid membranes, proteins, and biomolecular monolayers. The technique is sensitive to changes in molecular orientation, ordering, and interactions at these interfaces. In the context of lipid bilayers, for example, SFG can provide information into lipid packing, phase transitions, and the influence of biomolecules on membrane structure.

Moreover, SFG spectroscopy has been employed to investigate the adsorption and folding of proteins at interfaces. It can provide details about the secondary structure and conformational changes of proteins upon adsorption to surfaces. SFG spectroscopy is ideal for studying biomolecular interactions in aqueous environments. It can probe water structure and hydrogen bonding at interfaces, providing information on how biomolecules interact with water molecules. This information is vital for understanding hydration layers around biomolecules and the influence of water on biological processes. SFG spectroscopy plays a vital role for probing biological interfaces, providing unique insights into molecular structures and interactions.

Chapter 3

Methods and results

In this section, I show the effects of water around lipid monolayer systems, and within protein condensates and their surroundings. Both systems involve a balance of hydrophobic and hydrophilic forces at their interfaces, influencing macromolecular organization. Understanding these connections is critical for elucidating how water mediates biological functions in diverse contexts, from membrane-associated processes to the regulation of protein condensates, revealing the universal importance of water as a mediator of biomolecular interactions.

3.1 Water orientation around lipid monolayers

To investigate the effects of lipid head group chemistry, tail length, and tail saturation on water orientation around lipid monolayers, we simulated six different lipid monolayers. Three of these monolayers had PE headgroups with varying tail lengths and saturation. The remaining three monolayers had the same tail chemistry, but a PC headgroup. These lipid variations are depicted in figure 3.1.

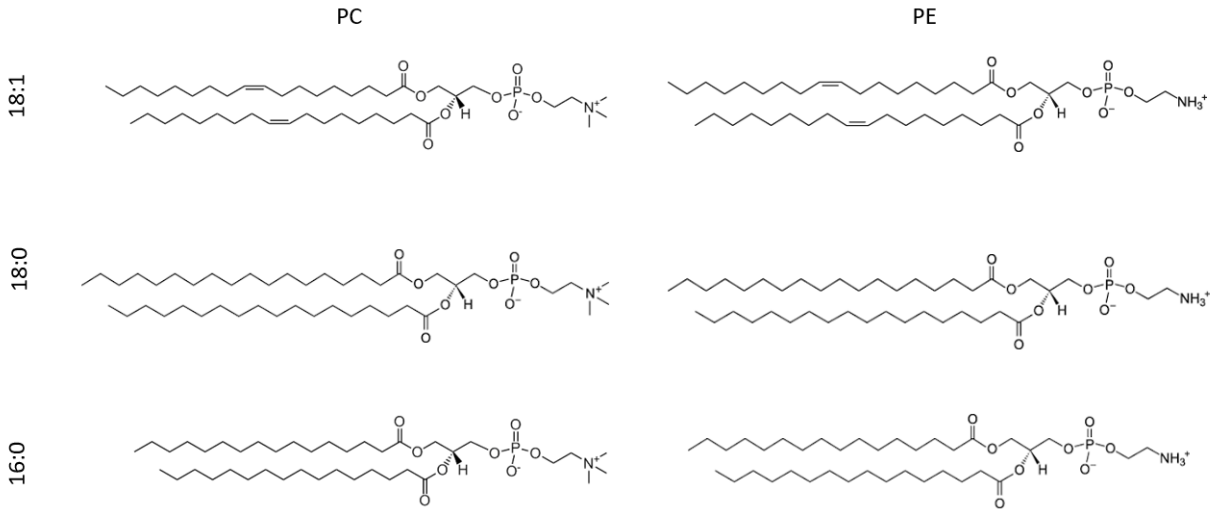

Figure 3.1: **Lipid chemistry.** The numbers on the left denote lipid tail length, followed by the number of saturated bonds. As an example, 18:1 denotes a tail length of eighteen carbons and one unsaturated bond. At the top, PC and PE represent phosphocholine and phosphatidylethanolamine headgroups, respectively.

The different chemistry of the lipids have an affect on the way how they interact among themselves, and with water molecules. A brief explanation on how the three main lipid components, headgroup, saturation and tail length, affect the lipid properties was given in the introduction section.

3.1.1 Simulations

Our systems consisted of a simulation box of 7 X 7 X 25.5 nm, with the elongated section in the Z direction, a water slab at the middle of the simulation box, and lipid monolayers surrounding the water slab, as shown in Figure 3.2. The systems contained ∼6552 water molecules and 150 mM of NaCl ions. Each leaflet contained ∼70 lipids.

The simulations were performed with the charmm36m forcefield, which is ideal for lipid simulations (*5*), and the TIP3 water model (*51*). The system was then equilibrated for 2 ns using the Berendsen (*52*) Barostat and thermostat. Afterwards, another 21 ns simulation was performed with the Nose-Hoover (*53*) thermostat at 300 K and the Parrinello-Rahman barostat (*54*) at 4.5e-5 bar^{-1}.

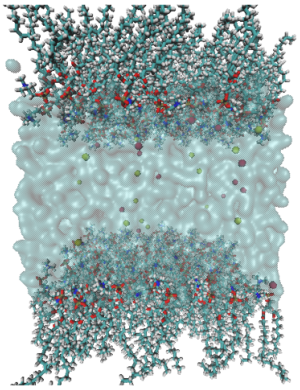

Figure 3.2: **Simulation box for the lipid monolayers.** The cyan transparent center represents the water molecules. The molecules on top and bottom of the water slab are lipids.

In order to understand the arrangement of water molecules at the lipid interface we calculate the tetrahedral order parameter of the molecules at a cutoff distance of 3.5 Å from the phosphate of the lipids.

This parameter measures the degree of tetrahedral arrangement of water molecules, reflecting their local structural order. Water naturally tends to form tetrahedral configurations due to the orientation of hydrogen atoms around the central oxygen atom, where each water molecule can be surrounded by four neighbors in a tetrahedral geometry.

At lipid monolayer interfaces, the tetrahedral order parameter provides insights into how water molecules organize. lipid interfaces disrupt the bulk tetrahedral arrangement due to interactions with the lipid headgroups. This metric helps to characterize the interfacial hydration structure, revealing how water molecules adapt to the environment at the monolayer interface.

To calculate the water tetrahedral order parameters we developed a code that searches a water molecule within a selected cut off radius, and then calculates the tetrahedral order parameter taking into account the nearest four water molecules, by using equation 3.1. We repeat this process until we have calculated the order parameter for every single water molecule in the hydration layer of the lipid monolayer. With this

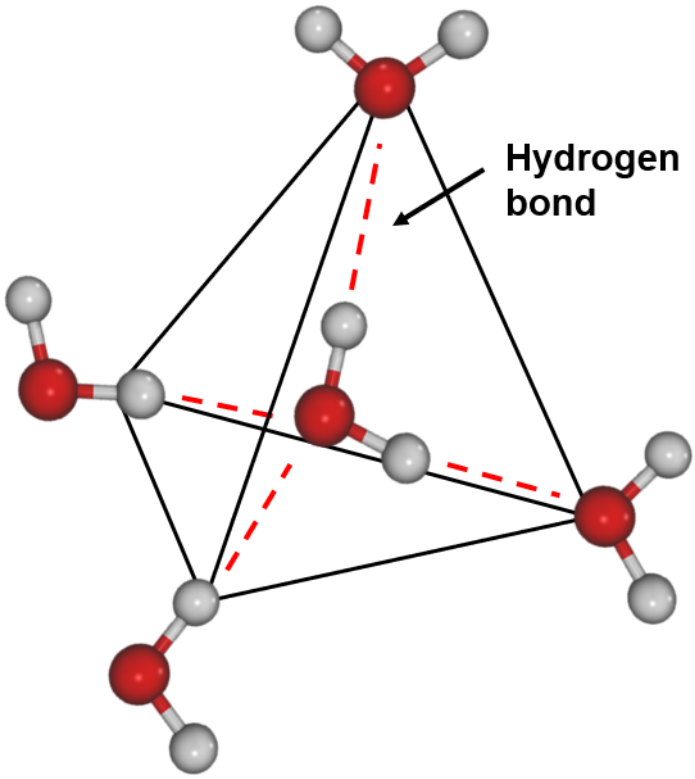

Figure 3.3: **The tetrahedral order parameter quantifies how closely the water molecules in a given environment adhere to this idealized tetrahedral structure.**
.

process we were able to calculate the order parameter of water all the water molecules at the interfae of the monolayers.

$$q = \frac{3}{2}\left(\frac{1}{2}\langle 3\cos^2(\theta) - 1\rangle\right) \tag{3.1}$$

in this equation q is the tetrahedral order parameter and θ represents the angle between the reference water molecule and its four closest neighbors, the closer the tetrahedral order parameter goes to 1 the closer the structure around a reference water molecule resembles an ideal tetrahedron.

As shown in figures 3.4 and 3.5, the tetrahedral order parameter distribution and mean values in both lipid types (PC and PE), are similar between the lipid types and consistent with the values for tetrahedral order parameter of water molecules at

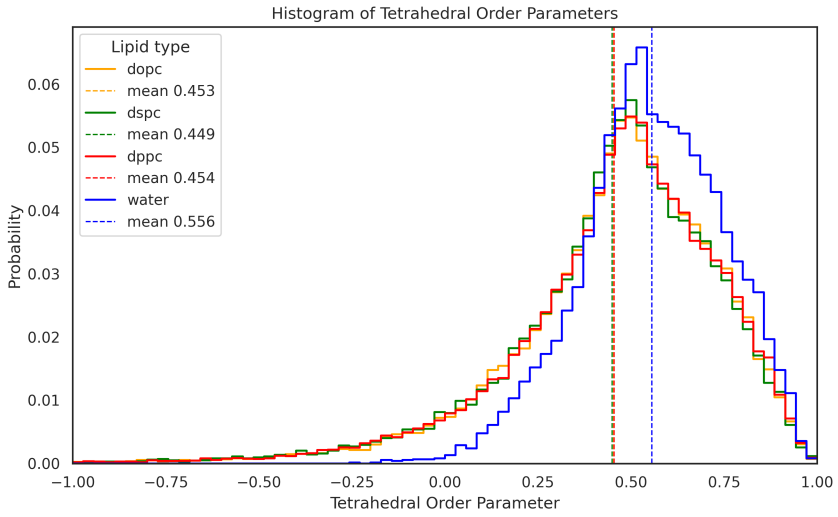

Figure 3.4: **Tetrahedral order parameter distribution** at the interface of PC lipid monolayers, the orange, green and red lines depict the distribution for dopc, dspc and dppc monolayer, the blue line depicts the distribution of bulk water, and the dotted colored lines display the mean values of the tetrahedral order parameter for its corresponding lipid

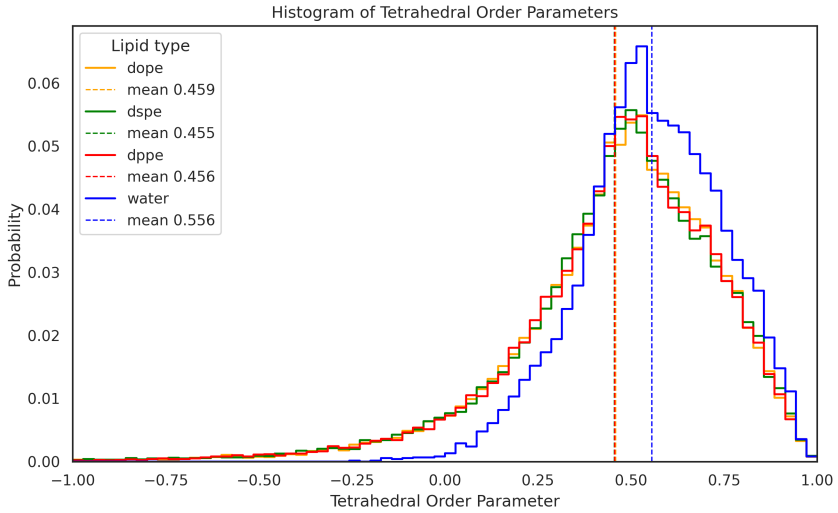

Figure 3.5: **Tetrahedral order parameter distribution** at the interface of PE lipid monolayers, the orange, green and red lines depict the distribution for dope, dspe and dppe monolayer, the blue line depicts the distribution of bulk water, and the dotted colored lines display the mean values of the tetrahedral order parameter for its corresponding lipid

other planar interfaces (*55*), greatly differing from an ideal tetrahedral water structure (order parameter value of 1), and with a difference of ∼0.1 from the tipical value for TIP3P bulk water (*56*). This effect is due to the fact that water molecules at the lipid-water interface are confined to a planar geometry, which limits their ability to form a complete tetrahedral structure. In bulk water, tetrahedral arrangement is possible because water molecules can freely interact in three dimensions. Near a plane,

water molecules lose one degree of freedom as their positions are constrained parallel to the surface. This confinement disrupts the three-dimensional hydrogen bond network between water molecules, forcing water to adopt a less ideal, more planar arrangement.

At the lipid interface, water molecules are often oriented in specific ways to maximize interaction with the surface. The oxygen atom of water may face toward the surface to act as a hydrogen bond acceptor, or the hydrogen atoms may orient toward the surface to donate hydrogen bonds.

To analyze the effect of different lipid chemistry on the orientation of water molecules, in relation to the monolayer, we calculated the angle between the vector of the dipole moment of the water molecule and the vector normal to the monolayer as shown as theta in figure 3.6.

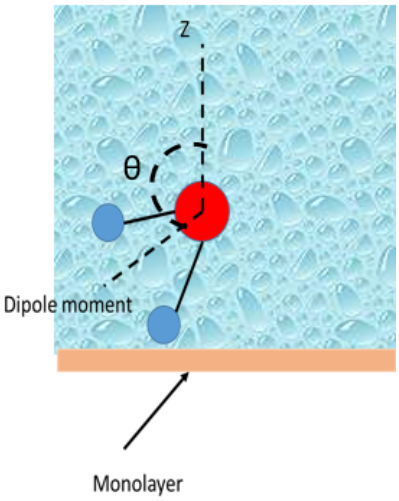

Figure 3.6: **Depiction of the vectors used as reference for calculating orientation of water molecules in reference to the membrane**. The light brown bar represents the surface of the monolayer, and the dotted lines show the vectors of the dipole moment and the normal to the surface of the monolayer, with theta depicting the angle between them.

The angle of all individual water molecules was measured and the mean angle of all the water molecules at different distances from the monolayer was used for the analysis. First we compared the water angle around lipids with same PC headroup, Figure 3.7

(left) and then we compared the water orientation whith PE lipide headgroups, Figure 3.7 (right).

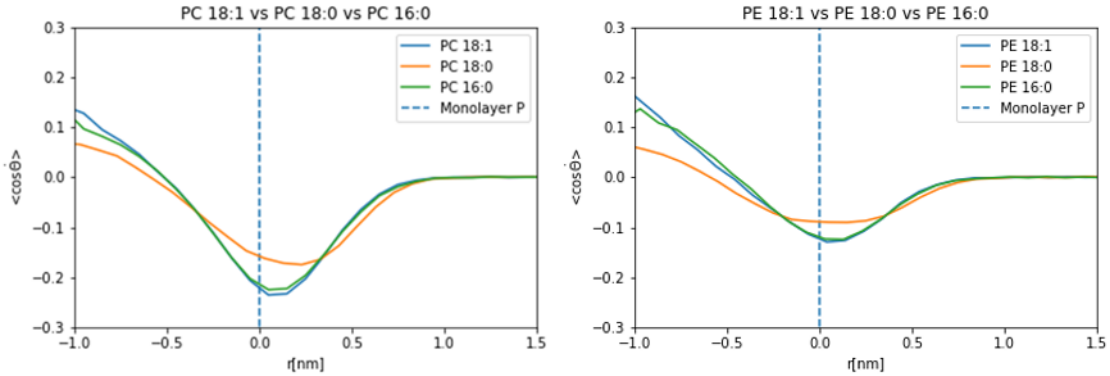

Figure 3.7: **Water orientation plot of water molecules in relation to monolayer surface** of the different lipid tail chemistries. The blue dotted line shows the position of the lipid monolayer. The colored lines indicate the different lipid chemistries. The X axis displays in nanometers the distance of the water molecules to the lipid monolayer, and the Y axis the mean (denoted by brackets) cosθ value of the water dipole angle to the monolayer. In this plot due to the topology of our simulation box, negative mean cosθ (closer to -1), shows that water molecules are, on average, oriented with their dipoles pointing toward the surface (hydrogen atoms close to the surface) of the monolayer. A cosθ of 0 indicates that water molecules, on average, have a random orientation relative to the surface, with no preferential alignment.

As shown in Figure 3.7, the overall behavior of water orientation between the PC and PE headgroups is similar, with the difference being that the PC lipids, with their larger hydrophilic groups have an increased effect on water orientation in the vicinity of the monolayer than their PE counterparts (cosθ value closer to -1). Nevertheless, we expected different water orientation regimes between the three lipid tail chemistries, (three separated colored lines). But is clear that we only find two regimes one for the 18:0 tails (long and saturated), and a different one for both the 18:1 (long with one unsaturation) and 16:0 (short and saturated).

This could be explained by the fact that introducing a saturation to the lipid tail induces a conformational change in the packing of the lipids equivalent to the change provoked by shortening the tail. To observe if this was indeed the case, we decided to plot the tilt angle of the lipids. The tilt angle of a lipid can be calculated by

determining the angle between the lipid tail vector (in this case drawn from the first to the last carbon of the tail) and the normal to the monolayer, as shown in figure 3.8.

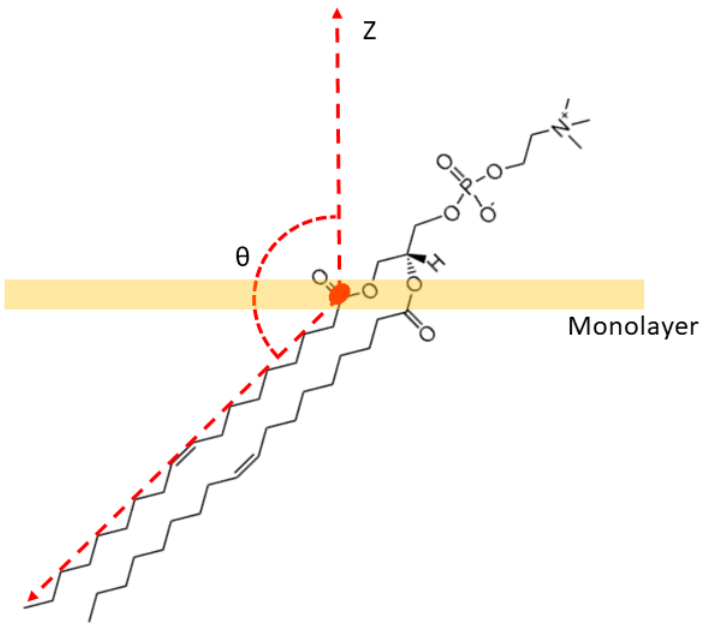

Figure 3.8: **Depiction of lipid tilt angle where the slab represents the monolayer.** The vector going down depicts the lipid tail vector, the Z vector depicts the normal to the monolayer and Θ represents the angle between them, also known as lipid tilt angle.

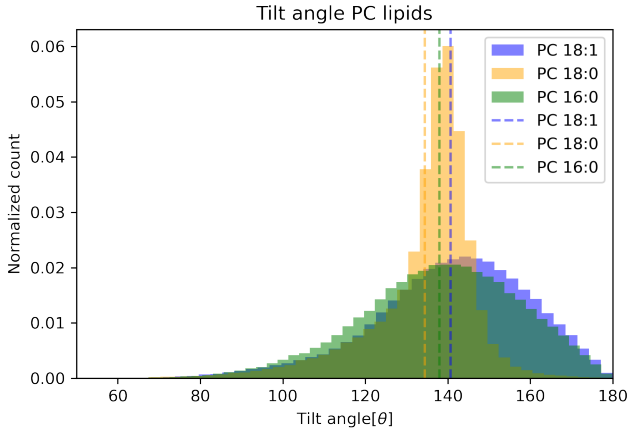

Figure 3.9: **Histogram of the normalized distribution of lipid tilt angles.** The colored areas represent the distribution for each lipid chemistry, the dashed line, the mean of their distributions

From figure 3.9 is noticeable that the 18:1 and 16:0 lipid tails are indeed packed in a similar fashion, which can be an explanation for the similarities in water orientation between them. In contrast the tails of the 18:1 lipid show a narrower distribution of tilt angles. which is expected from a fully saturated long lipid tail.

As shown in Figure 3.7, lipid tails with unsaturated bonds (18:1) have stronger effect on water orientation than the fully saturated lipid (18:0). An explanation for this is that unsaturated lipids are characterized by looser packing, which means that the water can create a larger hydration shell at the monolayer. In contrast, water molecules around lipids with fully saturated tails (18:0) will form a tight packing where water molecules have to compete with lipid-lipid interactions to interact with the polar headgroups.

3.2 Water structure and dynamics around protein condensates

To explore the influence of water dynamics in and around protein condensates, we adopted a multidisciplinary approach that integrates both experimental and computational techniques. Using spectroscopic methods, conducted by my colleague [redacted], alongside MD simulations conducted by me, we aimed to characterize the structure and dynamics of water molecules within protein FUS LC condensates. This combined methodology allows for a comprehensive understanding of hydration dynamics, enabling us to link experimental observations with detailed molecular insights and providing a robust framework for studying protein-water interactions in complex biological systems.

3.2.1 Slowing of water dynamics inside protein condensates

This section draws extensively from our published work. See Krevert and Chavez et al., 2023 (*57*).

To delve into the dynamics within biomolecular condensates (BCs), we focus on the low-complexity domain of FUS LC using IR spectroscopy. A comparative analysis is conducted between phase-separated FUS LC BCs at neutral pH 7.4 values and FUS LC at pH 11, which does not exhibit phase separation due to tyrosine deprotonation. Additionally, we examine a phosphomimetic mutant of FUS LC, denoted FUS LC 12E, at pH 7.4, shown in Figure 3.10, which remains uniformly distributed in solution and does not undergo liquid-liquid phase separation (LLPS) to form droplets, even at millimolar concentrations. Notably, FUS LC self-assembles into micrometer-sized BCs at neutral pH values, as expected, at a protein concentration of approximately 300 μM. Conversely, FUS LC 12E remains uniformly dispersed even at a protein concentration of about 400 μM.

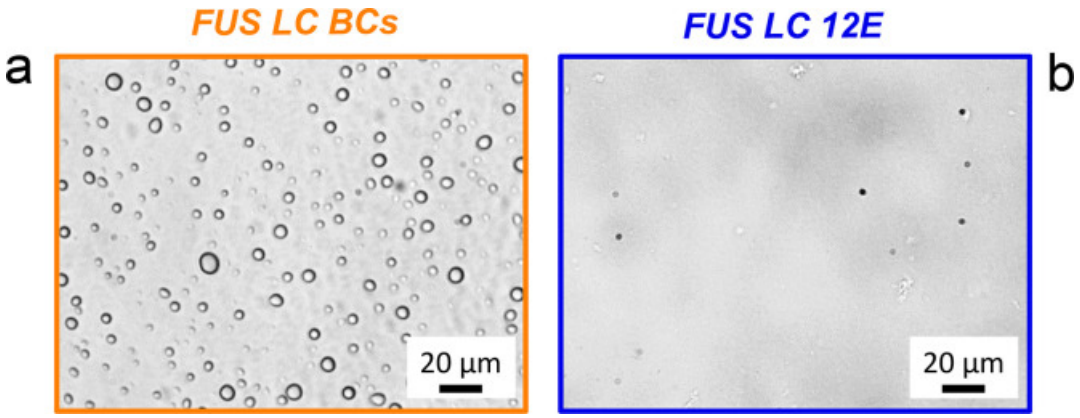

Figure 3.10: **Phase contrast microscopy captured images** of (a) FUS LC BCs exhibiting a diameter range of 2–10 μm, and (b) the homogeneous distribution of FUS LC 12E. These images were acquired following 2D IR experiments. Notably, panel (b) displays an absence of droplets. Figure and text taken from (*57*)

.

The amide I mode of proteins, associated with the C=O stretching mode of the protein backbone, serves as a highly informative indicator of alterations in both protein secondary structure and its surrounding environment. Hence, to investigate spectral variations during LLPS, we examine the amide I vibrational band of FUS LC for both its homogeneous and condensed states. The linear Fourier transform IR spectrum shown in Figure 3.11 illustrates the profiles of FUS LC BCs at pH 7.4 shown in orange, FUS LC 12E at pH 7.4, shown with the blue line, and FUS LC at pH 11 shown in green.

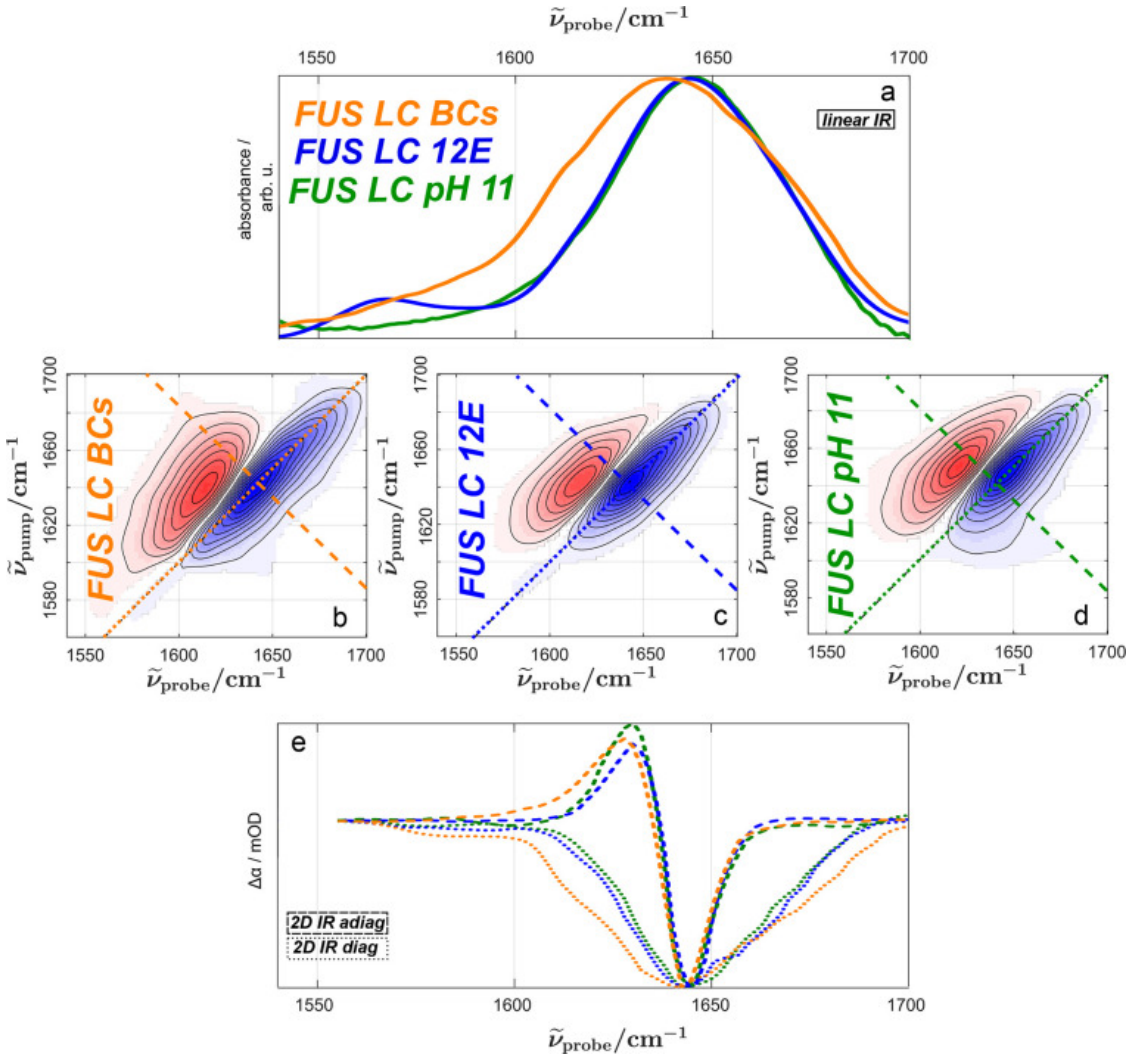

Figure 3.11: **Linear IR spectra of FUS LC under different conditions** (a) spectra of FUS LC BCs in orange, FUS LC 12E at pH 7.4 in blue, and FUS LC at pH 11 in green. The concentrations for FUS LC at pH 11 and FUS 12E were maintained at approximately 0.4 mM. All spectra underwent correction for solvent background and were normalized to the absorption of the amide I mode at around 1640 cm–1. (b), (c), and (d) present 2D IR spectra at a 0 fs waiting time for FUS LC BCs, FUS LC 12E, and FUS LC at pH 11, respectively. The dotted lines signify the diagonals where the pump frequency equals the probe frequency, while the dashed line represents an antidiagonal at 1645 cm–1. (e) showcases diagonal and antidiagonal cuts through the 2D IR spectra, as depicted in panels (b–d). Figure and text taken from (*57*)

..

The amide I band observed in the infrared spectra of FUS LC BCs, FUS LC 12E, and FUS LC at pH 11 exhibits notable breadth. In all samples, the amide I band centers around ∼1640 cm–1. Despite this commonality, distinctions in the band shape are discernible between homogeneous forms (FUS LC 12E and FUS LC at pH 11) and phase-separated proteins (FUS LC BCs). In the condensed state, the amide I band of

FUS LC appears somewhat broader, featuring weak shoulders at approximately ∼1580 and ∼1620 cm–1, distinguishing it from the amide I peak of FUS LC 12E and FUS LC at pH 11. Consequently, the broadened infrared spectra imply that the amide backbone of FUS perceives a more diverse distribution of microenvironments within the BCs, compared to the uniformly dissolved protein. The slight red-shift potentially indicates stronger hydrogen bonds of the amide backbone with water and/or proteins. In order to further confirm this interpretation, we calculated the IR spectra from all-atom molecular dynamics simulations shown in Figure 3.12. These calculations substantiate the idea that spectral alterations primarily arise from a distinct hydrogen-bonding environment of the protein within the BCs, in contrast to that of the homogeneously dissolved protein.

Vibrational spectroscopy and NMR, provide insights into the structural dynamics of the FUS LC protein in its condensed versus non-condensed states. Vibrational spectroscopy, which captures rapid structural and environmental changes on a very short timescale (subpicosecond), shows minor differences in FUS LC's vibrational structure between these two states. On the other hand, NMR, which averages structural information over a much longer timescale (milliseconds), suggests that FUS LC does not undergo significant structural changes when it condenses (*59*).

This discrepancy in findings raises an important question about whether the structural differences observed with vibrational spectroscopy are transient or if they might persist longer but simply aren't detectable on the NMR timescale.

To address this question, we conducted 2D IR spectroscopy experiments on the amide I band (*60*). In a typical 2D IR experiment, vibrational modes undergo excitation with intense femtosecond IR laser pulses, and the response of the excited sample is probed using an infrared probe pulse. The response is recorded as a function of excitation frequency, enabling the detection of the system's response to a specific pump frequency. This approach facilitates the experimental determination of frequency–frequency correlations (FFCs) for molecular-level oscillators. By adjusting the

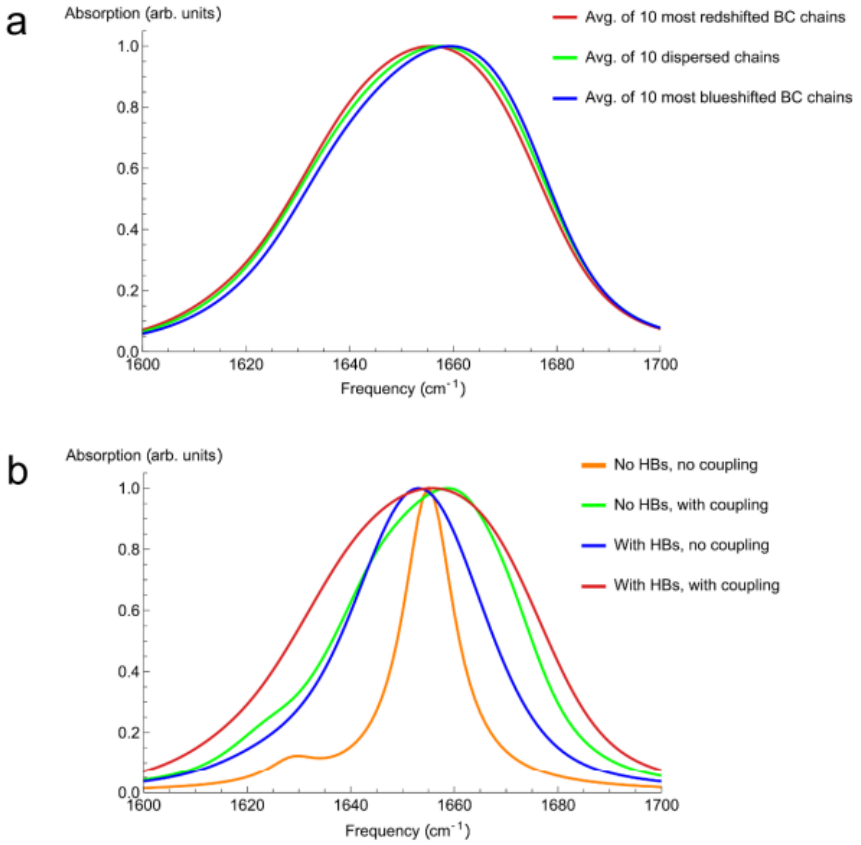

Figure 3.12: **Spectral calculations from MD dynamics** reveal a broad distribution within the ensemble of BC chains, with the observed spectral broadening predominantly influenced by hydrogen-bonding effects. a) Within the diverse ensemble of protein chains residing in the BC, distinct sub-ensembles exhibit either red- or blue-shifted characteristics (illustrated by the red and blue traces) in comparison to dispersed chains (depicted by the green trace). The limited extent of red- and blue-shifting may be attributed in part to the presence of a relatively small number of chains within the droplets, with a substantial fraction likely located on the condensate surface. b) While both coupling and hydrogen bonding contribute to spectral broadening, it appears that relatively strong hydrogen bonding plays a dominant role in shaping the spectra of the 10 most redshifted FUS chains. This determination is based on the systematic activation and deactivation of different spectral modeling elements governing the spectra, calculated using the one-exciton Hamiltonian approach In the absence of any coupling or hydrogen-bonding model, the local modes are fixed to the gas phase frequency (*58*). Amide groups downstream of proline residues, which feature a heavier ring moiety instead of the proton found in other amino acids at the nitrogen atom, are redshifted by 19 cm-1 as per (*58*). When coupling models are introduced, the normal modes broaden as anticipated, yet the predominant red-shifting results from the inclusion of the hydrogen-bonding model. The combined influence of both effects leads to even greater spectral broadening, aligning with expectations in these one-exciton Hamiltonian calculations. Figure and text taken from (*57*)

.

delay between the probing pulse and the excitation pulses, referred to as waiting time (T_w), is possible to investigate vibrational dynamics.

Panels b–d in Figure 3.11 show the 2D IR spectra for FUS LC BCs (3.11b), for homogeneous solvated FUS LC 12E (3.11c), and FUS LC pH 11 (3.11d) at a T_w of 0 fs. The shown spectra display the typical 2D IR features of an inhomogeneously broadened band, a ground state bleach (blue shaded areas) at the excitation frequency ($\nu_{pump} = \nu_{probe}$), and a red-shifted ($\nu_{pump} > \nu_probe$) excited state absorption ($1 \rightarrow 2$ transition) of the anharmonic oscillators (red shaded areas). The homogeneous line width, which can be assessed from the antidiagonal cut of the 2D spectra shown with dashed lines in Figure 3.11 panels b–d, is comparable at 1645 cm–1 for the condensed and homogeneous form of FUS LC (dashed lines in (3.11e)). Conversely, the diagonal cuts dotted lines shown in panel e of the same figure resemble linear IR spectra. Given the different sensitivities of the linear (squared transition dipole) and 2D IR spectroscopy, the similarity of the squared linear spectra and the diagonal cuts indicates that the distribution of transition dipoles is similar for FUS LC in the homogeneous and condensed phases (*50*). The broadening of the amide I band upon phase separation is reflected in both the linear and the diagonal cuts of 2D IR spectra. Together, these results suggest that upon BC formation, the intrinsic properties of the amide I mode (the homogeneous line widths in Figure 3.11d) are unaffected. However, the distribution of microenvironments of the amide groups is somewhat altered inhomogeneous line width, (diagonal cut shown as dotted line in Figure 3.11d). In general, the resonance frequency of an amide mode is sensitive to the H-bonding environment and concomitantly to the secondary structure of a protein (*50*, *61–63*). As such, the enhanced spectral inhomogeneity of FUS LC in the droplets may stem from different protein conformations, resulting in different distributions of inter and intramolecular protein interactions (*64*). However, due to the reduced volume density of solvating water molecules, a broader distribution of the protein hydration motifs, which in turn alters the amide I frequency (*65*), may also give rise to spectral inhomogeneity.

The spectral inhomogeneity is investigated further to understand the underlying reasons behind the spectral variation in the amide I band. This exploration delves into the vibrational dynamics of the amide mode of FUS, specifically analyzing the

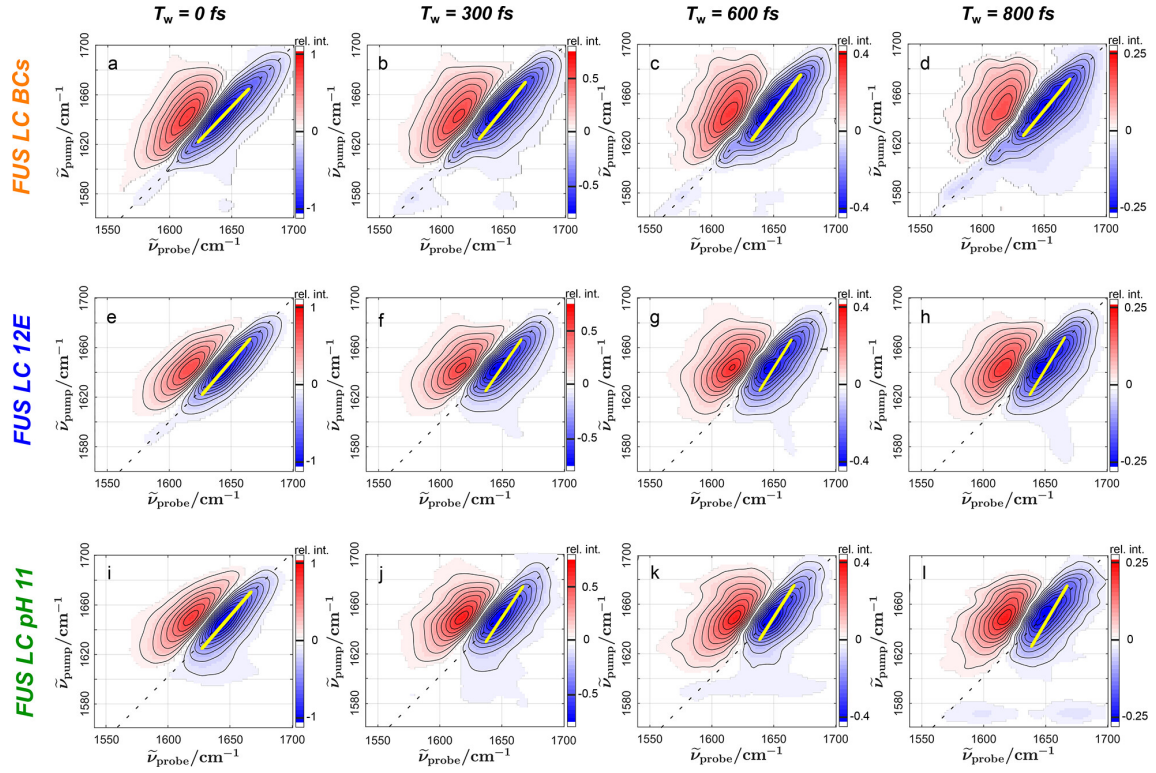

Figure 3.13: **BCs (top row, a–d), FUS LC 12E (center row, e–h), and FUS LC pH 11 (bottom row, i–l), with increasing T_w (from left to right).** Yellow lines show the center line slope (CLS) of the ground state bleach (blue). With an increasing T_w, the CLS, which is almost parallel to the diagonal when $T_w = 0$ fs, becomes more parallel to the pump axis. This loss of FFC is more pronounced for FUS LC 12E and pH 11 than for FUS LC BCs; spectral diffusion occurs faster for the homogeneous FUS LC than for the droplets. Figure and text taken from (*57*)

.

time-resolved 2D IR spectra for FUS LC 12E, FUS LC BCs, and FUS LC pH 11, as presented in Figure 3.13. The conformational and hydration dynamics, occurring on distinct time scales, are examined through the decay of transient signals with increasing T_w, from the maximum normalized bleaching signal in Figure 3.13a a to 25% of the original signal in Figure 3.13d. This decay is attributed to the relaxation of the excited state population to the vibrational ground state (τ_{VER}). The relaxation is quantified by fitting a single-exponential decay [A_0 exp(-T_w/τ_{VER}) + y_0] to the integrated peak volumes, revealing similar vibrational lifetimes (τ_{VER}) of 0.54 ± 0.03 ps for FUS LC 12E, 0.55 ± 0.03 ps for FUS LC BCs, and 0.48 ± 0.02 ps for FUS LC pH 11. These values closely resemble the relaxation of the amide group of the dipeptide alanyl-alanine in D_2O (0.7 ps) (*66*). The observed τ_{VER} values align with those found for other proteins (∼1 ps or less) (*50*, *67*) and the isolated amide moiety of N-methylacetamide

(0.5 ps) (*68*). Therefore, the study indicates that condensation does not significantly alter these dynamics, consistent with the general insensitivity of energy transfer from excited amide groups to the primary structure of proteins.

Despite the comparable vibrational energy relaxation observed in both the homogeneous and condensed phases of FUS LC, significant disparities emerge in the decay of FFCs, representing the temporal variations in the instantaneous frequency of the amide I modes. The assessment of FFCs can be done qualitatively by examining the elongation of signals along the diagonal in the 2D spectra. In all three samples, the 2D IR peaks at a (T_w) of 0 fs exhibit noticeable elongation along the diagonal due to the inhomogeneous broadening of the amide I mode. As we increase the T_w to 800 fs in 3.13, these FFCs become less pronounced, with the peaks adopting a more vertical orientation.

However, in the case of BCs, the ground state bleaching signal retains some elongation along the diagonal of Figure 3.13d. In contrast, for solutions of FUS LC 12E and FUS LC at pH 11, the peaks have already shifted to a more vertical alignment (Figure 3.13h,l. Consequently, the detected probe signals exhibit a weaker correlation to the excitation frequency at a T_w of 800 fs for homogeneous protein solutions compared to the condensed state.

Center line slope (CLS) is a common metric for quantifying FFCs. CLS is calculated as the inverse slope of a straight line fitted through the minima of transient signals (in the ground state bleach) corresponding to a specific pump frequency (depicted by yellow lines in Figure 3.13). In the case of a perfect direct correlation between excitation and detection frequencies, the CLS = 1, while a CLS of 0 signifies that the detected response is independent of the excitation frequency (indicating a line parallel to the pump axis). Figure 3.14 presents these CLS values over T_w. The figure illustrates that the initial CLS is somewhat higher for the condensate compared to the homogeneous FUS LC from either FUS 12E or FUS LC at pH 11, indicating increased amide I mode spectral inhomogeneity in the condensate. It's worth noting

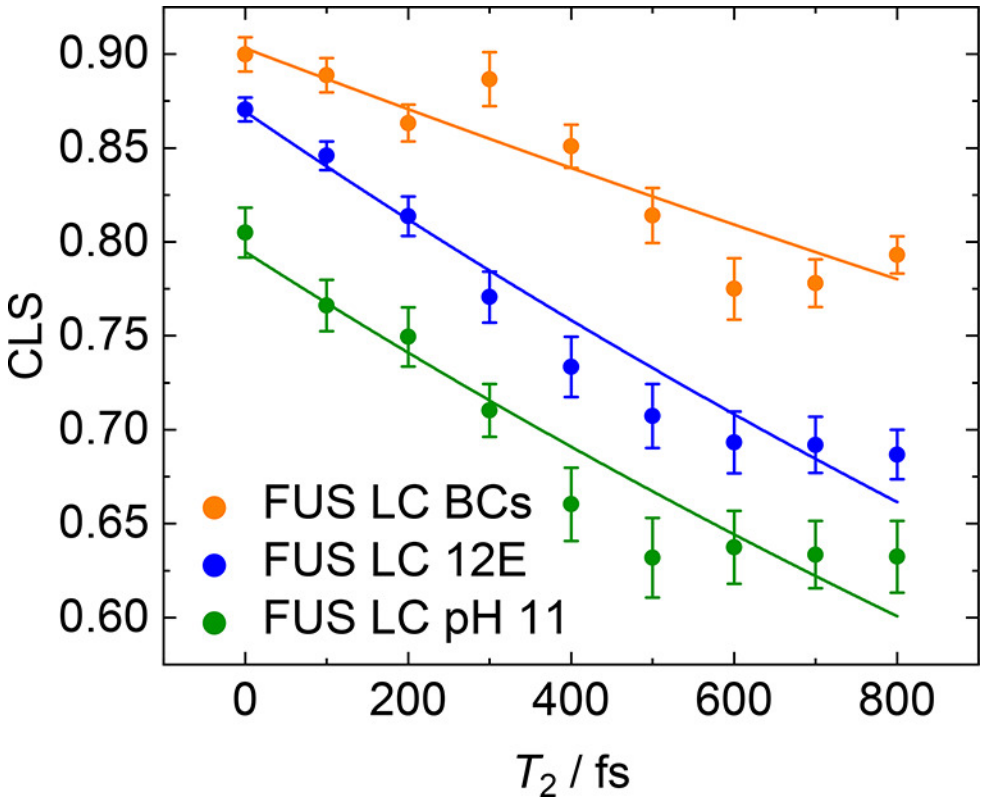

Figure 3.14: **(T_w)-dependent CLS of FUS LC BCs (orange), FUS LC 12E (blue), and FUS LC pH 11 (green).** The CLS of BCs exhibits a greater value for instantaneous inhomogeneous broadening [CLS(T_w = 0 fs)] when compared to homogeneous FUS LC in a solution. Moreover, the CLS of BCs demonstrates a slower decay than that of FUS LC in homogeneous solutions. Furthermore, the decay rate of the CLS in BCs is slower compared to that of FUS LC in homogeneous solutions. Single-exponential fits are depicted by solid lines. The determination of CLS for all three samples utilized the same frequency range (1618–1685 cm–1). Notably, even with a significant variation in this frequency range, these trends remain consistent. Figure and text taken from(*57*)

.

that the starting CLS value is higher in FUS 12E (at pH 7.4) than in FUS LC at pH 11, implying distinct amide I spectral inhomogeneity under conditions where FUS is homogeneous. This observation aligns with the protein's disordered nature and adaptability in various solution environments(*69*). Additionally, considering the nodal line slope yields similar results. We observed that an alternative consideration of the nodal line slope yields comparable outcomes. Besides the distinct initial (T_w = 0) inhomogeneity, there are notable differences in spectral dynamics among various protein states. Within the experimentally observable time frame of approximately 800 fs, the CLS value experiences a decay of about 20% for homogeneous FUS LC (FUS LC 12E and FUS LC at pH 11). In contrast, in BCs, the CLS value decreases by

around 10%, indicating a twofold slower dynamics of the amide mode in the condensed phase compared to the homogeneous FUS LC. To characterize the decay kinetics, we employed an exponential decay model [CLS(T_w) = CLS_0 exp(-$T_w/\tau_{specdiff}$), depicted by solid lines in Figure 3.14]. This enabled the determination of decay times ($\tau_{specdiff}$) under the assumption of a straightforward single-exponential decay. Despite the fact that the experimental CLS does not completely decay within 800 fs, suggesting the potential presence of multiple time scales in the decay dynamics, the short vibrational lifetime imposes constraints for employing more intricate models(*70*). Nonetheless, the use of single-exponential fits seems to effectively capture the experimental data shown in Figure 3.14, allowing for the evaluation of differences in spectral diffusion dynamics between the two samples. From these analyses, we determine that the diffusion time $\tau_{specdiff}$ for FUS LC 12E is 2.9 ± 0.2 ps (2.9 ± 0.3 ps for FUS LC at pH 11), which is nearly twice as fast as the value observed for FUS LC BCs ($\tau_{specdiff}$ = 5.5 ± 0.7 ps). The fitted results corroborate the approximately 2-fold slower spectral diffusion dynamics of FUS LC within condensed droplets compared to its behavior in the homogeneous solution. The decay of the CLS can be attributed to protein fluctuations (conformational dynamics) and/or environmental fluctuations (H-bonding with water in the proteins hydration shell), both of which can influence instantaneous amide I frequencies(*50*, *71–73*). While NMR relaxation rates have indicated slowed backbone dynamics in the condensates on the pico to nanosecond timescale(*74*), we observe significant differences in subpicosecond spectral dynamics for FUS LC in condensed and homogeneous environments. These subpicosecond dynamics may be influenced by diverse protein structures, such as a sub-ensemble of aggregate proteins associated with the shoulder at 1620 cm–1. However, the observed deceleration remains consistent across the frequency range of the CLS analysis. Alternatively, distinct energy transfer dynamics may impact spectral dynamics(*60*), yet the similarity in vibrational energy relaxation time scales in all three samples suggests minimal changes in energy dynamics. Furthermore, the observed picosecond time scales align with H-bond dynamics, where fluctuations and formation/dissociation events can lead to a loss of FFCs, given the sensitivity of amide resonance frequency to hydration(*71*). Consequently, we posit

that the pronounced differences in CLS decay dynamics shown in Figure 3.14 primarily stem from a deceleration of the hydration dynamics of FUS LC within the BCs.

In order to test this hypothesis, we compared our experimental observations with dynamics derived from fully atomistic MD simulations of a FUS LC droplet immersed in water. Due to the large size of the system, it was necessary to start from a coarse grained conformation, provided by Benayad *et al.* (*75*). This CG condensed conformation was performed by utilizing a modified MARTINI CG model (*76*), and the simulation trajectory file was supplied by the authors of this study. Subsequently, the system underwent backmapping using CHARMM-GUI (*77*) and the CHARMM-GUI Martini Maker (*78*), which employs the backward.py program for the all-atom conversion (*79*).

The backed-mapped all-atom system consists of 134 chains representing the low-complexity region of the FUS LC. These chains are situated within an approximately 50 nm cubic box and are explicitly surrounded by 4,222,084 water molecules and 24,202 Na and Cl ions, leading to a salt concentration of approximately 150 mM. As a result, the simulation box encompasses a total of 17,213,858 atoms.

The all atom simulation utilized the GROMACS 2019.3 simulation suite(*43*, *44*), employing the a99SB-disp force field(*80*) for protein interactions and the four-point a99SB-disp(*81*) for water molecules. The equations of motion were resolved with a time step of 2 fs, and periodic boundaries were applied in all directions. Covalently bonded hydrogens were constrained using the P-LINCS(*82*) algorithm. Long-range electrostatic interactions were computed using particle-mesh Ewald summation(*81*), while Lennard-Jones and short-range electrostatic interactions were truncated at a cutoff distance of 12 Å.

For the equilibration, we used a multiple step equilibration process. The initial step involved energy minimization of the system using the steepest descent algorithm. Following this, a series of equilibration simulations were performed with position restraints applied to the heavy atoms of the FUS chains. These steps included:

1. Two 125 ps NVT simulations with a 1 fs time step, using the Berendsen thermostat (coupling constant of 1 ps) and restraint force constants of 1000 and 400 kJ mol^{-1} nm^{-2}.

2. A 125 ps NPT simulation with a 1 fs time step, the Berendsen thermostat and barostat (coupling constants of 1 ps and 5 ps, respectively), and a restraint force constant of 200 kJ mol^{-1} nm^{-2}.

3. Two 500 ps NPT simulations with a 2 fs time step, the Berendsen thermostat and barostat (coupling constants of 1 ps and 5 ps, respectively), and restraint force constants of 200 and 100 kJ mol^{-1} nm^{-2}, respectively.

4. A 1.5 ns NPT simulation using the same parameters as in step 3 but without position restraints.

After this, the system underwent an equilibration period of 20 ns NPT ensemble at 300 K, using the Par- rinello–Rahman barostat (time constant of 2 ps) and the Bussi–Donadio–Parrinello velocity rescaling (V-rescale) thermostat (coupling time constant of 0.1 ps). to allow for the relaxation of the system in accordance with the atomistic model. With the system relaxed we did a production run of 1 ns that we utilized for the following analysis.

We clasified each protein in the simulation as "dispersed (dilute FUS LC)" or "condensed (FUS BCs)" by using the gmx cluster analysis function within the Gromacs simulation package, employing a cutoff distance of 0.3 nm. This categorization was further confirmed using MDAnalysis(*83*, *84*), ensuring that proteins categorized as dispersed had no contact within 0.3 nm of any protein categorized as condensed.

After the classification we proceeded to evaluate the dynamics of hydration, by examining the lifetime of hydrogen bonds formed between the amide groups of protein chains and water molecules. The criteria to classify if a hydrogen bond is occurring between a water molecule and the amide group is depicted in Figure 3.15. With this information we calculate the survival function, which is essentially the time correlation function shown in equation 3.2 of the hydrogen bonds. (*85*).

$$C(t) = \langle \frac{\sum(h_{ij}(t_0)(t_0+1))}{\sum(h_{ij}(t_0)^2)} \rangle \tag{3.2}$$

Here, the function h_{ij}(t) is defined as 1 when there exists a hydrogen bond between the amide oxygen of residue i and any hydrogen of water molecule j at time t, otherwise, it is set to 0. The angular brackets indicate the averaging across various starting points of the trajectory.

We fit the hydrogen bond time autocorrelation function by eliminating the baseline H-bond population, and then fitting with a triple-exponential function shown in equation 3.3 (*85*).

$$f(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} + A_3 e^{-t/\tau_3} \tag{3.3}$$

in equation 3.3 the pre-exponential factors are limited to the condition shown in equation 3.4.

$$A_n = 1 - \sum_{i=1}^{n-1} A_i \tag{3.4}$$

The resulting time autocorrelation functions are shown in 3.16. These findings indicate that the overall hydration dynamics exhibit qualitative similarities between FUS LC in condensed and dilute phases. There is an initial rapid decay of C(t) followed by a somewhat slower decline in the H-bond correlation function. However, at t >

	τ (ps)	τ_1 (ps)	τ_2 (ps)	τ_3 (fs)	A_1	A_2	A_3
condensed	3.2 ± 0.2	7.5 ± 0.3	0.77 ± 0.13	50.0 ± 0.1	0.41 ± 0.01	0.07 ± 0.01	0.52 ± 0.02
in solution	2.3 ± 0.2	5.4 ± 0.2	0.58 ± 0.15	50.0 ± 0.2	0.41 ± 0.01	0.07 ± 0.01	0.52 ± 0.02

Table 3.1: A_j and τ_j are the amplitudes and time constants, respectively, of the three exponentials. τ is the amplitude-weighted average time constant. Errors were obtained from the diagonal elements of the covariance matrix of the nonlinear fit. Figure and text taken from(*57*)

0.5 ps, it becomes evident from the raw data that C(t) decays noticeably more slowly for FUS LC within the condensates than for dilute FUS LC, indicating prolonged protein–water H-bond lifetimes in the condensates.

$$\tau = \int_0^\infty C(t)\, dt \tag{3.5}$$

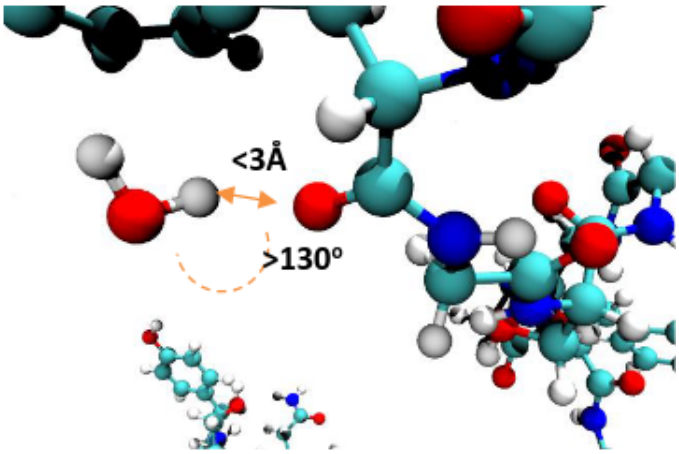

Figure 3.15: **Depiction of water molecule forming a hydrogen bond with FUS** A hydrogen bond is identified between a water molecule and the amide I oxygen if any hydrogen atom of the water molecule is within a 3 Å distance from the amide I oxygen and exhibits an angle greater than 130 degrees in the O(water)-H(water)-O(FUS amide) configuration. Figure taken from(*57*)

.

The amplitudes and time constants for the fittings are shown in table 3.1. The swiftest dynamics, characterized by a time constant of approximately 50 fs, exhibit comparability between the protein states in both solution and the condensate. These dynamics can be associated with rapid water motions, such as the librational dynamics of water hydrogen bonds, although numerical effects in the analysis may also contribute. However, the slower dynamics, characterized by time constants of approximately 0.5–0.8

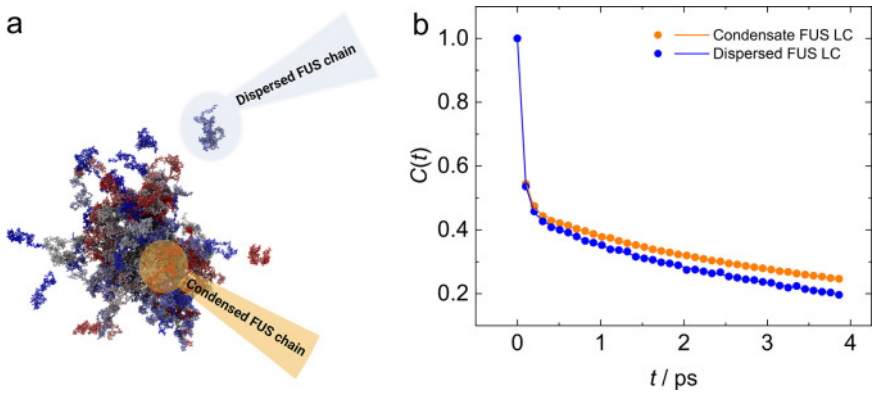

Figure 3.16: **(a) Depiction of the categorization of individual FUS protein chains in the atomistic simulations.** A straightforward clustering analysis was conducted to recognize components of the protein droplet, identifying groups of chains in close proximity. The remaining protein chains are designated as proteins in the dispersed phase. (b) H-bond autocorrelation function between the carbonyl oxygen and water for FUS LC BC (orange) and FUS LC dilute (blue), revealing the decelerated hydration dynamics in the condensed phase. The solid lines represent the triple-exponential fit, and the corresponding fit parameters are outlined in Table **??**. Figure and text taken from(*57*)

.

and 5–7 ps, consistently display a slower value in the condensate compared to FUS LC in solution. Both time scales align with the experimental 2D IR results and represent typical time frames for H-bond lifetimes.

The observation of two distinct time scales in the simulations (0.5–0.8 and 5–7 ps) suggests potential heterogeneous water accessibilities to the amide backbone. However, the longer time scales (5–7 ps) could also be attributed to H-bond breaking induced by the motion of multiple heavy atoms, such as side chain rotation. Nonetheless, this analysis confirms a significant extension (by approximately 40%) in the lifetimes of H-bonds between water and the amide C=O group for FUS LC in the condensates compared to dilute FUS LC. Considering the intimate relationship between the hydration state of the amide C=O and the amide I resonance frequency, this observed increase in H-bond lifetimes in the simulations provides a substantial explanation for the experimentally observed slower decay of FFCs in the 2D IR experiments.

In essence, we investigated the impact of FUS LC's LLPS on the vibrational dynamics of the amide I mode, a distinctive vibrational indicator for the protein backbone, utilizing time-resolved vibrational spectroscopy. Linear infrared spectra unveil a 12%

broader line width for FUS LC within the condensate compared to FUS LC at pH 11 or FUS LC 12E at neutral pH, both of which remain completely homogeneous as solutions. Robust FFC signals in 2D IR spectra indicate that the line widths observed in linear IR spectra stem from inhomogeneous broadening, implying a more extensive range of microenvironments for the amide backbone of FUS LC in the condensates compared to the diluted FUS LC.

3.2.2 Water structure around FUS LC protein

In the previous section we analyzed water dynamics in the whole FUS LC condensate and diluted FUS LC protein chains. In this section we aim to understand the ordering of water levels at an aminoacid level resolution. In order to achieve this we used the data of the MD simulations from the previous section.

In the lipid monolayer section, we used the water tetrahedral order parameter to analyze the organization of water molecules at lipid interfaces. This parameter quantified how water adapts to the structural and chemical environments presented by lipid monolayers. At these interfaces, the disruption of the bulk tetrahedral arrangement due to interactions with lipid headgroups and tails provided insights into interfacial hydration structure and dynamics.

Now, we extend the use of this parameter to study water organization in FUS LC systems. Specifically, the tetrahedral order parameter is applied to investigate the structural arrangement of water molecules within the hydration shell of FUS LC chains. To calculate the order parameter, we used the same developed code that identifies water molecules within a cutoff radius of 3.5 Å from any FUS LC chain. With this methodology we are able to assess the order parameter for all water molecules around FUS LC..

This parameter provides insights into the structural organization of water molecules, particularly within confined spaces like protein condensates. This parameter aids in understanding the hydration properties of biological macromolecules, as well as the

interactions and structural characteristics of water in the specific microenvironments, within around FUS LC chains.

In order to compare the ordering of water around FUS LC, with water in bulk state, we performed a small 5x5x5 nm simulation with the same parameters of the FUS LC simulation, consisting of only water molecules, and then calculated the order parameter for every water molecule in the box.

With this approach, we were able to calculate the water tetrahedral order parameter of water around protein chains in condensed, diluted, and bulk state. The result of this analysis is shown in Figure 3.17. Notably, this parameter is similar for waters inside the protein and around diluted protein chains. Nevertheless, in both systems the order parameter value is lower than water in bulk state.

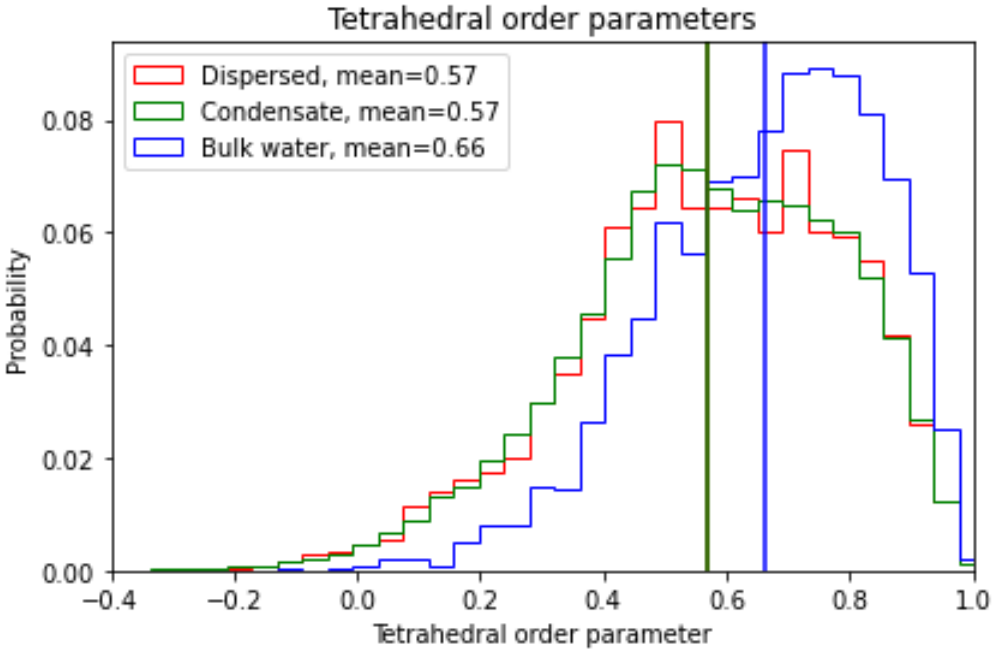

Figure 3.17: **Tetrahedral order parameter distribution** in red, water molecules around dispersed protein, in green water molecules around protein in condense state, and in blue the order parameter of bulk water. The vertical straight lines denote the mean tetrahedral order parameterl

.

3.2.3 Water structure in protein condensate microenvironments

To delve deeper into the water structure within protein condensates, we analyzed the water environment in distinct microenvironments of the condensate. Specifically, our goal was to comprehend the water structure around particular amino acid types or specific amino acid contacts.

Due to the vast number of particles in the system, we opted to utilize graph theory methods, in order to simplify and organize our data structures as shown in Figure 2.2. In doing so, we created a graph network where each node contained attributes with information of their amino acid type, which FUS LC protein they belonged to, or if they where a water molecule. If two particles were within the specified cutoff, we added an edge between the nodes to represent a contact between them. To obtain information on the distance between the nodes, we assigned a weight to the edges based on the proximity of the particles. This weight can range from 0.01 to 1, respectively indicating the maximum and closest distances between the nodes, respectively.

Using this methodology we first look into the contact map between the distinct amino acids of the protein chains shown in figure 3.18. Of interest is that Tyrosine (Y) and Glutamine (Q) have an increased number of contacts, suggesting that this aminoacids could be responsible of stabilizing the condensate by interacting with neighboring protein chains (*59*).

We scrutinized the water order parameters encompassing the specific amino acids within the protein chain. As depicted in Figure 3.19, the mean tetrahedral parameter of the water surrounding these amino acids is notably similar. However, the distribution of the order parameter for water molecules around the amino acids identified as high-contact residues, is more sharply defined than the ones around their low-interacting counterparts. This shows that the water surrounding the high-contact residues has fewer potential tetrahedral conformations compared to their counterparts.

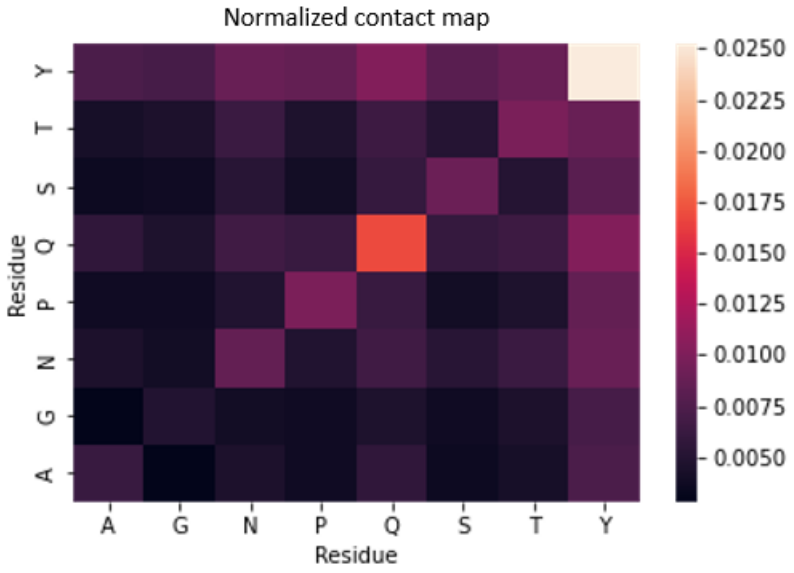

Figure 3.18: **Contact map** between FUS LC condensate aminoacids, normalized by their abundancy in the sequence.

We expanded our analysis by creating density plots that compare the degree of individual amino acids to the proportion of isolated water molecules in their vicinity. A water molecule is considered isolated if it has fewer than four neighboring water molecules within a 3.5 Å cutoff radius. These plots provide insights into the relationship between the narrow distribution of the tetrahedral order parameter around high-contact amino acids and their interaction patterns within their local microenvironments.

Our previous findings showed that water surrounding high-contact residues exhibits a more sharply defined tetrahedral order parameter compared to low-contact residues. These density plots extend this analysis by focusing on the connectivity of the amino acids themselves. The degree reflects the number of interactions an amino acid forms within the molecular network, highlighting its role in maintaining structural cohesion. Meanwhile, the proportion of isolated water molecules indicates the extent to which water is confined in small pockets or trapped between neighboring FUS LC chains, where it can act as a stabilizing element in the network, such as by forming hydrogen-bond bridges.

By plotting these variables together, we assess how the interaction levels (degree) of amino acids correlate with the presence of isolated water molecules in their environment. The results show that high-contact residues not only exhibit greater connectivity within

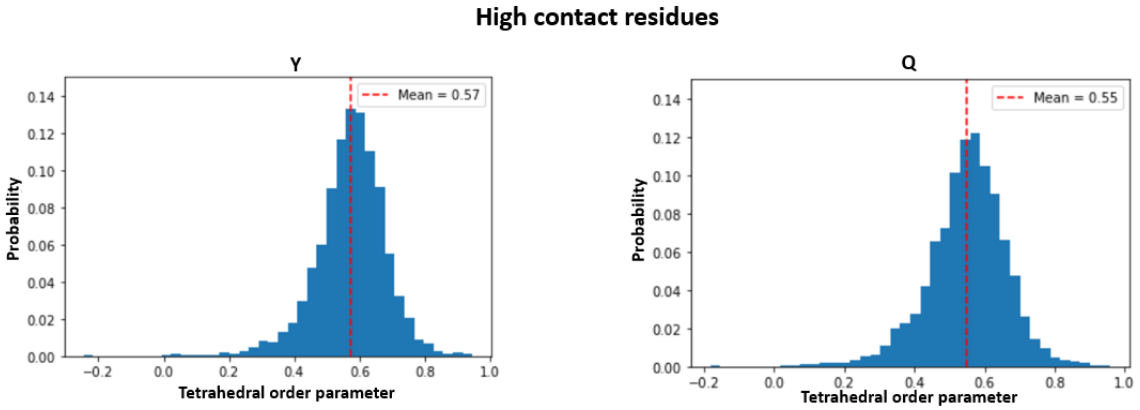

Figure 3.19: **Histograms of tetrahedral order parameter distribution** around residue types of FUS LC, the red dotted lines display the mean values of the tetrahedral order parameter.

the network but also interact more frequently with isolated water molecules. Conversely, low-contact residues, which correspond to broader distributions of tetrahedral order, are less likely to engage in such interactions with isolated waters, making them less involved in network stabilization.

The plots specifically reveal that high-contact residues such as glutamine (Q) and tyrosine (Y) have steeper slopes compared to other residues. This indicates that when their surrounding microenvironment consists predominantly of isolated water molecules, these residues form highly dense contact networks. This suggests that isolated water molecules play a critical role in stabilizing these networks, acting as structural bridges that enhance the cohesion and stability of the system.

These results show that although the water dynamics differ between FUS LC BC and dispersed FUS LC, the tetrahedral structure of water is largely conserved in both systems, nevertheless it seems that there are certain restrictions in the values that the order parameter can take at the micro environments around high contact residues, which could be explained by the amount of isolated waters around this residues.

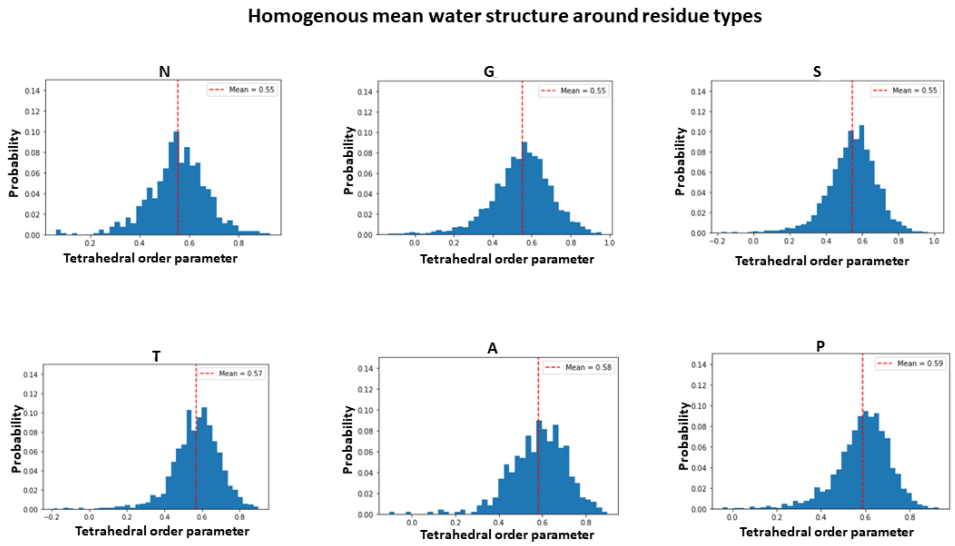

Figure 3.19: **Histograms of tetrahedral order parameter distribution (Continuation)** around residue types of FUS LC, the red dotted lines display the mean values of the tetrahedral order parameter.

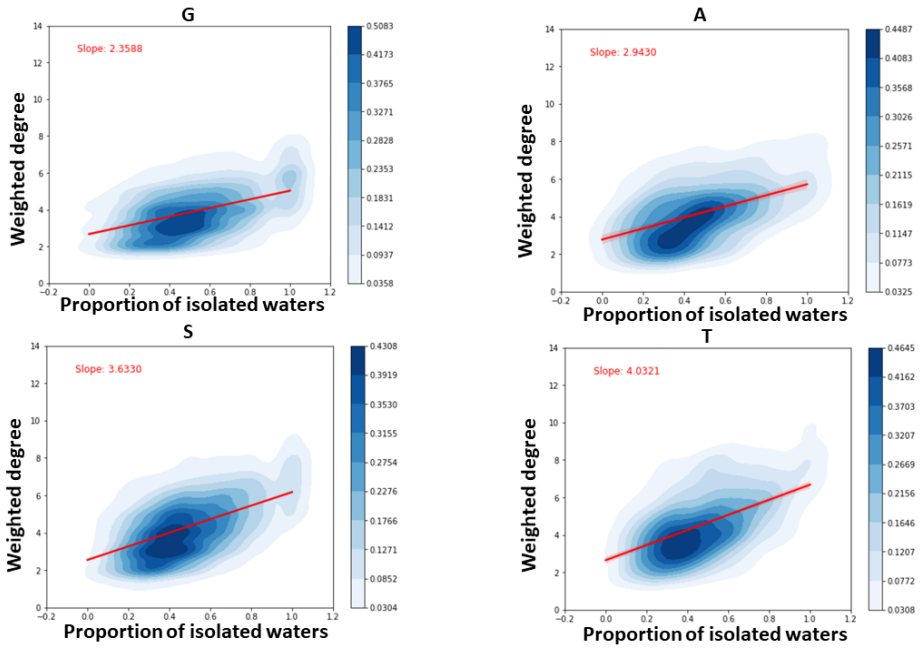

Figure 3.20: **Density map plot of aminoacid degree** in relation to the proportion of isolated waters by aminoacid type arranged by aminoacid type from top left to bottom right. The Y axis displays the weighted degree (amino acid contacts) and the X axis shows the proportion of Isolated waters (less than 4 water molecule neighbors). The numbers in red denote the value for the slope of the linear fitting shown as a red line.

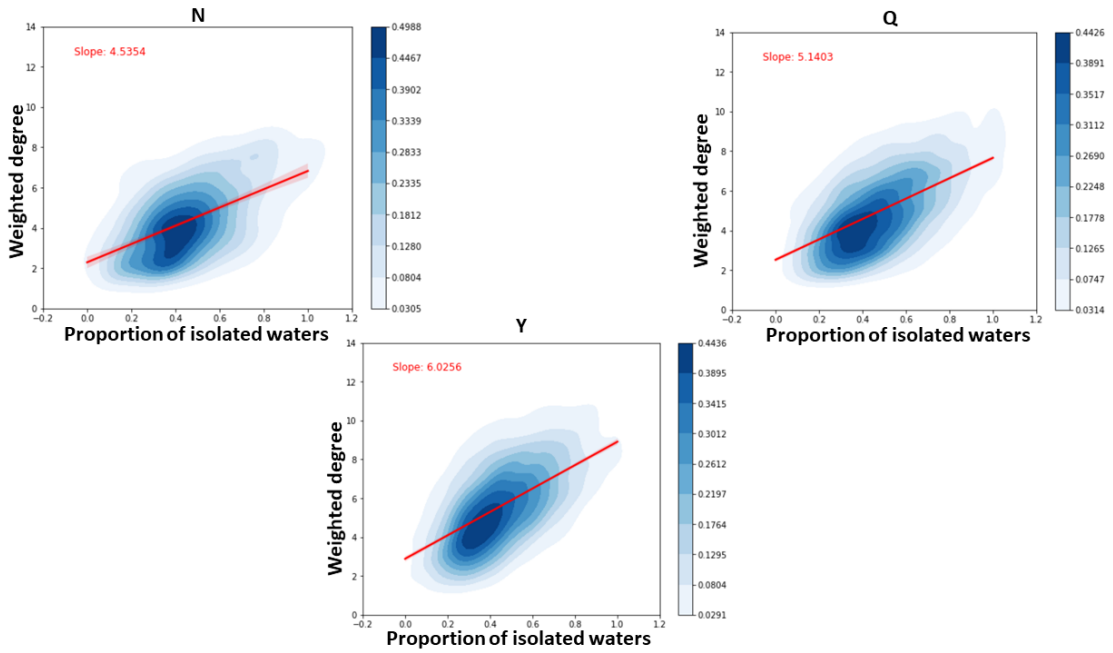

Figure 3.20: **Density map plot of aminoacid degree (continuation)** in relation to the proportion of isolated waters by aminoacid type arranged by aminoacid type from top left to bottom right. The Y axis displays the weighted degree (amino acid contacts) and the X axis shows the proportion of Isolated waters (less than 4 water molecule neighbors). The numbers in red denote the value for the slope of the linear fitting shown as a red line.

Chapter 4

Discussion

In this project, we used computational methods to characterize the structure and dynamics of water molecules interacting with lipids and protein condensates. By employing MD simulations, and spectroscopy techniques, we were able to investigate water organization, hydrogen-bonding dynamics, orientation around lipid interfaces and water structure within protein condensates. These analyses provided insights into how water structure and dynamics differ within phase-separated protein structures. Our approach also enabled the identification of distinct water orientations across different lipid packing arrangements, and specific protein-protein contact points within condensates, revealing how water structure and dynamics are altered around these biological structures.

4.1 Lipid Organization and Water orientation

In the first part of this dissertation we aimed to understand whether water orientation is affected by lipid tail or headgroup chemistry. In order to achieve this, we performed MD simulations of six different lipid monolayer systems with varied specifications. We focused on two headgroups, PC and PE, fully saturated or with one unsaturation, and two different tail lenghts, 16 and 18 carbons. Based on this systems, we calculated the orientation of the water molecules at the lipid-water interface. From

the results shown in Figure 3.7, we highlight that, as expected, hydrophilic interactions between the lipid headgroup and water molecules, enhances the orientation affinity of water molecules to the lipid monolayer. Nevertheless due to the similitude between the 16:0 and 18:1lipids, the water orientation plots (Figure 3.6) show no discernible water orientation differences. This reveals that, in this case, lipid tail chemistry didn't affect water orientation. By analyzing lipid tilt angles we discovered that water orientation is highly dependant of lipid packing, measured through tilt angles. From this we can conclude that the determining factor affecting water orientation at the lipid-water interface lipid packing in the monolayer.

4.2 Slowing of water dynamics inside FUS LC protein condensate

Given the relevance that LLPS proteins play in RNA transcription and neuron motor diseases, and the fact that these biological systems are mostly composed of water, we aimed to understand the differences in the dynamics of water inside protein condensates and diluted proteins. Due to the fact that FUS LC domains drive phase separation and that FUS condensates play a role in RNA transcription and DNA repair, FUS LC is an ideal model for studying LLPS in proteins. To achieve this, we studied a system comprising both, FUS LC condensates and diluted FUS LC proteins shown in Figure 3.10.

A highlight of our methodology is the focus on experimental collaboration. Our colleagues at the Spectroscopy Department at the MPIP carried out a spectroscopy experiment indicating that water dynamics are approximately twice as slow inside FUS LC BC in comparison to diluted FUS LC (Figure 3.14). In order to interpret these experimental findings at an atomistic resolution, we utilized all-atom molecular dynamics simulations.

From our H-bond time autocorrelation plot (Figure 3.16b), the slowing of water dynamics inside FUS LC BCs is apparent. Furthermore, by fitting the time autocorrelation function to the triple exponential function (3.3), we discerned three different time constants associated with bond dynamics time scales. We associate the fastest constant to the librational mode of water, the middle time scale to hydrogen bonding dynamics, and the slowest to conformational changes in the protein backbone. From the fitting function, we derived an amplitude-weighted average time constant τ, shown in Table 3.1, that clearly presents a 40 % increase in hydrogen bond lifetime in FUS LC BC, compared to the diluted FUS LC. Our data shows that the dynamics are consistently slower in the condensate compared to FUS LC in solution. This hints at the notion that slowed hydrogen-bonding dynamics contribute to the unique solvation and reaction properties observed within the condensate.

Moreover, our results revealed a contrast with hydrogels, where macroscopic transport and microscopic hydration dynamics show a correlation for FUS (**Roget**). This correlation can be rationalized by considering that the H-bond lifetimes in the hydration shells are closely linked to exchange dynamics. The reduction in the density of water molecules within the condensates leads to a decrease in H-bond exchange rates within the BCs, resulting in more persistent hydration shells around FUS and potentially other solutes in the BCs. This suggests that the unique hydration dynamics within FUS condensates may play a critical role in modulating the stability and biochemical environment of FUS LC condensates.

4.3 Water structure around FUS LC protein

In order to understand water organization around FUS LC at an amino acid resolution, we used a combination of graph analysis techniques combined with calculation of tetrahedral order parameter of water molecules from MD simulations. The tetrahedral order parameter allowed us to measure water structure around FUS LC proteins as it quantifies how closely the local arrangement of water molecules resembles a tetrahedral structure. In this way, we objectively quantified how closely the water organization

around FUS LC resembles the structure of bulk water. Through this analysis, we assessed the changes in water organization near the FUS LC-water interface. Observing the plot from Figure (3.17), we see a clear difference between the mean tetrahedral order parameter of TIP4P bulk water (0.66), and around FUS LC proteins (0.57), both in condensed and diluted phase. The difference between bulk water and FUS LC systems is of 0.09, which could look negligible at first sight. Nevertheless, it is important to consider that an increase mean value of 0.06 is the difference between room temperature water and water below freezing point (**duboue**). Such a difference in mean values highlights the magnitude of water structure disordering in the neighborhood of the FUS LC protein. This could be explained by the fact that around FUS LC, water molecules find themselves in a crowded micro environment due to inter/intra-protein contacts severely restricting tetrahedral conformation.

Likewise, it is possible that that water molecules are taking part in forming hydrogen bond bridges between amino acids. When we calculate the order parameter of a water molecule that is forming a hydrogen bond with FUS LC, we can expect neighborhood of water molecules that don't resemble a tetrahedron. In our analysis, we also examined the tetrahedral order parameter of water molecules surrounding each amino acid within the FUS LC protein chains. Shown in Figure 3.9, the mean tetrahedral order parameter of water around the amino acids is generally similar across different residue types. However, we observed that the distribution of the order parameter is more narrow for water near residues (Y) and (Q), detected as high-contact residues in the contact map of Figure 3.18, compared to their low-contact counterparts. This suggests that water molecules near high-contact residues experience a more restricted range of possible tetrahedral conformations, indicating a localized influence of protein interactions on water structuring.

These findings demonstrate that while the water dynamics differ between FUS LC condensates and dispersed FUS LC, the tetrahedral organization of water is generally preserved in both environments. Nonetheless, water molecules in the immediate vicinity of high-contact residues show limited available conformations in tetrahedral ordering,

likely due to specific microenvironmental restrictions imposed to these protein-water interactions. This implies an interaction landscape where specific residue contacts influence local water structuring without fully disrupting the tetrahedral order of the surrounding hydration shell.

Chapter 5

Conclusion

5.1 Biological implications

In the aforementioned projects, we used the combined advantages of MD, 2D IR and graph analysis to elucidate the intricate interactions between water-lipid and water-protein biological systems. Trough our work, we showed that water is a dynamic component in biological systems, and a fundamental component of protein condensates.

In our studies of water orientation around lipid monolayers, we found that the main factor in orienting the water molecules was the lipid tilt angles, closely related to lipid packing.

The hydration layer around lipids affects the membrane's elasticity and bending rigidity. These properties are fundamental for processes like vesicle formation, membrane fusion, and cell motility, understanding how water molecules organize themselves at the lipid-water interface is indispensable to develop new therapeutic treatments, a recent example is that of the new RNA vaccines, which require lipid vesicles for the delivery of mRNA, acting as protective and delivery vehicles that ensure the introduction of mRNA into human cells. As nanoparticles, they are carefully designed to interact with cell membranes, allowing the mRNA to enter the cells. Understanding

water organization around these vesicle may open the door to improve vesicle stability and vesicle-cell interaction.

Furthermore, in this work we explored the dynamics of water within condensates. As mentioned in previous sections protein condensates play a vital role in the realm of biology, influencing the spatial arrangement of various components or altering hydrodynamic transport.

FUS LC is a great study target within this field. There have been extensive studies on how FUS condensates concentrate proteins and RNA, in order to organize and regulate biochemical reactions. Inside FUS droplets, FUS binds to specific RNA molecules, creating a local environment rich in RNA. This helps control RNA processing, splicing, and transport. Nevertheless the hydrogen-bonding environment within these condensates, a crucial factor in solvation, was not previously understood.

Our study sheds light into the role of water-protein hydrogen-bonding dynamics within condensates, providing information on how these molecular interactions may underpin the specific solvation and reaction behaviors exhibited by these condensed biological structures.

Finally, our MD studies provide insight into the structural organization of water molecules, particularly within confined spaces like protein condensates. Our results aid in understanding the hydration properties of biological macromolecules, as well as the interactions and structural characteristics of water within specific cellular environments. In this case, the tetrahedral order parameter serves as a valuable tool for analyzing and characterizing the water structure under different contexts contexts, especially in the study of protein condensates and cellular processes.

In general, our results indicate that both the structure and dynamics of water exhibit distinct behaviors inside protein condensates compared to bulk water. This divergence in water behavior may serve as one of the driving mechanisms through which

protein condensates create a unique environment for facilitating biological processes within the cell.

5.2 Future research directions

5.2.1 Lipid organization and water orientation

Lipid tilt angles are closely related to lipid phase, which refers to the distinct structural and physical arrangements that lipids can adopt:

1. Gel Phase: Lipid molecules are tightly packed, with ordered, extended acyl chains (tails) and minimal movement.
2. Liquid-Ordered Phase: Lipids are organized, though some degree of lateral mobility is retained.
3. Liquid-Disordered Phase: Lipids have greater freedom of movement, with more rotation and lateral diffusion.

To understand water organization around lipids, it is essential to compare membranes in these different phases. One approach involves simulating various lipid chemistries across phases. However, we found it challenging to compare different lipids in the same phase while maintaining consistent thermodynamic properties; measuring water orientation at different temperatures hindered direct comparisons. To address this, we attempted simulations under uniform thermodynamic parameters, but encountered phase coexistence within different regions of the lipid monolayers, pushing for classification of each lipid by phase, and an independent analysis of water molecules surrounding each type of lipid phase.

In practice, our current algorithm to measure lipid tilt angles, could be used to classify lipids into specific phases. Additionally, a modified version of the algorithm we used to calculate water tetrahedral order parameter could be employed to measure water orientations around lipids within each distinct phase.

5.2.2 Water structure and dynamics around protein condensates

While this study provides insights into the organization and dynamics of water around FUS LC condensates, there are still several interesting research paths for further exploration. One direction could be to investigate how specific amino acid sequences within FUS LC influence local hydration dynamics. Certain residues, such as glutamine and tyrosine, are known to play a stabilizing role in phase separation, therefore, targeted studies could help us elucidate their specific effect in water dynamics, and, potentially, whether they alter protein interactions within FUS LC condensates.

Expanding this research work to study temperature and concentration effects on FUS LC hydration dynamics could also provide valuable insights. For instance, performing MD simulations across a range of temperatures and protein concentrations could clarify the conditions under which water structure and dynamics are disrupted or preserved the most around the protein. Understanding these conditions would be of great benefit, as they could have important implications in FUS LC stress responses, or in pathological aggregation linked to neurodegenerative diseases.

When it comes to water structure however, it is crucial to study not only water organization around different amino acid types but also around specific interactions between these residues. Particularly, examining water structuring around contacts like Q-Q, Y-Y, or Q-Y is of great interest due to their central role in stabilizing FUS LC protein condensates.

A current limitation in MD is managing the large datasets generated by our simulations. Even with graph analysis formalism, performing these analyses requires frequent application of clustering, parallelization, and classification algorithms to analyze each parameter effectively. This could be solved by the application of machine learning techniques to analyze the large datasets from MD simulations. By employing machine learning models, it may be possible to identify subtle patterns and structural characteristics in water-protein interactions that are not apparent through traditional

analysis methods. Such approaches could reveal relationships between hydration behavior and phase separation tendencies, which may be crucial for understanding the stability of FUS LC condensates, and, ultimately, lead to the development of therapeutic strategies aimed at modulating FUS LC condensate stability and mitigating aggregation-related diseases.

Acknowledgements

First, I express my deepest gratitude to my supervisor, Prof. [redacted], for his invaluable guidance, insights, encouragement, and patience throughout this journey. I am equally grateful to my direct supervisor, [redacted], whose expertise and constructive feedback provided me with the tools and abilities necessary to complete this project. Without his guidance, finishing this work would not have been possible. I also acknowledge my direct supervisor, [redacted], for his role in overseeing this project. Who helped in dealing with the many challenges that arose during the course of the project, his support encouraged me to approach the work with greater thoroughness and attention to detail.

I owe a special thanks to the Max Planck Institute of polymer research, particularly to the theory department including the administrative staff and my fellow colleagues, for providing an environment that fostered quality research and for the camaraderie that made this journey memorable.

To my peers and friends, [redacted], [redacted], [redacted] and [redacted]. Thank you for the amazing times we spent together and for always being there throughout the highs and lows of this journey. Your support and shared laughters were invaluable during my time at the MPIP.

I am profoundly grateful to my family for their unconditional love and encouragement. To my parents, [redacted] and [redacted], thank you for always having my back and for your unwavering support.

Finally, to my wife [redacted], your patience, love, and understanding have been my anchor through this challenging period of my life. I couldn't have done it without you. This thesis is dedicated to all those who have played a role, big or small, in helping me achieve this milestone. Thank you all from the bottom of my heart.

Bibliography

(*1*) Liu, Z., and Barrett, E. J. (2002). Human protein metabolism: its measurement and regulation. *American Journal of Physiology-Endocrinology and Metabolism 283*, Publisher: American Physiological Society, E1105–E1112.

(*2*) Tanford, C., and Reynolds, J., *Nature's Robots: A History of Proteins*, Google-Books-ID: Vg5H7InfABEC; OUP Oxford: 2003; 500 pp.

(*3*) Crick, F. (1970). Central Dogma of Molecular Biology. *Nature 227*, Publisher: Nature Publishing Group, 561–563.

(*4*) McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature 267*, Publisher: Nature Publishing Group, 585–590.

(*5*) Yu, Y., Krämer, A., Venable, R. M., Brooks, B. R., Klauda, J. B., and Pastor, R. W. (2021). CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *Journal of Chemical Theory and Computation 17*, Publisher: American Chemical Society, 1581–1595.

(*6*) CASE, D. A., CHEATHAM, T. E., DARDEN, T., GOHLKE, H., LUO, R., MERZ, K. M., ONUFRIEV, A., SIMMERLING, C., WANG, B., and WOODS, R. J. (2005). The Amber Biomolecular Simulation Programs. *Journal of computational chemistry 26*, 1668–1688.

(*7*) Shukla, D., Schneider, C. P., and Trout, B. L. (2011). Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. *Advanced Drug Delivery Reviews 63*, 1074–1085.

(*8*) Doan, L. C., Dahanayake, J. N., Mitchell-Koch, K. R., Singh, A. K., and Vinh, N. Q. (2022). Probing Adaptation of Hydration and Protein Dynamics to Temperature. *ACS Omega 7*, Publisher: American Chemical Society, 22020–22031.

(*9*) Grigera, J. R., and McCarthy, A. N. (2010). The Behavior of the Hydrophobic Effect under Pressure and Protein Denaturation. *Biophysical Journal 98*, 1626–1631.

(*10*) Dumetz, A. C., Chockla, A. M., Kaler, E. W., and Lenhoff, A. M. (2008). Effects of pH on protein–protein interactions and implications for protein phase behavior. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1784*, 600–610.

(*11*) Martini, S., Bonechi, C., Foletti, A., and Rossi, C. (2013). Water-Protein Interactions: The Secret of Protein Dynamics. *The Scientific World Journal 2013*, 138916.

(*12*) Nakasako, M. (2004). Water-protein interactions from high-resolution protein crystallography. *Philosophical Transactions of the Royal Society B: Biological Sciences 359*, 1191–1206.

(*13*) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. In *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, Pullman, B., Ed.; The Jerusalem Symposia on Quantum Chemistry and Biochemistry; Springer Netherlands: Dordrecht, 1981, pp 331–342.

(*14*) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. In *Molecular Biology of the Cell. 4th edition*; Garland Science: 2002.

(*15*) Ali, A., and Bagchi, A. (2015). An Overview of Protein-Protein Interaction. *Current Chemical Biology 9*, 53–65.

(*16*) Ryan, D. P., and Matthews, J. M. (2005). Protein–protein interactions in human disease. *Current Opinion in Structural Biology 15*, 441–446.

(*17*) Fry, D. C. (2006). Protein–protein interactions as targets for small molecule drug discovery. *Peptide Science 84*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.20608, 535–552.

(*18*) Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., and Shi, J. (2020). Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy 5*, Publisher: Nature Publishing Group, 1–23.

(*19*) Fox, P. F., Uniacke-Lowe, T., McSweeney, P. L. H., and O'Mahony, J. A. In *Dairy Chemistry and Biochemistry*, Fox, P. F., Uniacke-Lowe, T., McSweeney, P. L. H., and O'Mahony, J. A., Eds.; Springer International Publishing: Cham, 2015, pp 145–239.

(*20*) Dill, K., and Bromberg, S., *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*, 2nd ed.; Garland Science: New York, 2010; 1568 pp.

(*21*) Pocker, Y. (2000). Water in enzyme reactions: biophysical aspects of hydration-dehydration processes. *Cellular and Molecular Life Sciences CMLS 57*, 1008–1017.

(*22*) In *Wikipedia*, Page Version ID: 1245562117, 2024.

(*23*) Phospholipid — Cell Membrane, Lipid Bilayer & Fatty Acids — Britannica `https://www.britannica.com/science/phospholipid` (accessed Oct. 2, 2024).

(*24*) van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology 9*, Publisher: Nature Publishing Group, 112–124.

(*25*) Marsh, D. (1996). Lateral pressure in membranes. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes 1286*, 183–223.

(*26*) Nagle, J. F., and Tristram-Nagle, S. (2000). Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes 1469*, 159–195.

(*27*) Saliba, A.-E., Vonkova, I., and Gavin, A.-C. (2015). The systematic analysis of protein–lipid interactions comes of age. *Nature Reviews Molecular Cell Biology 16*, Number: 12 Publisher: Nature Publishing Group, 753–761.

(*28*) Gaber, M., Medhat, W., Hany, M., Saher, N., Fang, J.-Y., and Elzoghby, A. (2017). Protein-lipid nanohybrids as emerging platforms for drug and gene delivery: Challenges and outcomes. *Journal of Controlled Release 254*, 75–91.

(*29*) Landolt-Marticorena, C., Williams, K. A., Deber, C. M., and Reithmeier, R. A. F. (1993). Non-random Distribution of Amino Acids in the Transmembrane Segments of Human Type I Single Span Membrane Proteins. *Journal of Molecular Biology 229*, 602–608.

(*30*) Belmonte, S. L., and Blaxall, B. C. (2012). Conducting the G-protein Coupled Receptor (GPCR) Signaling Symphony in Cardiovascular Diseases: New Therapeutic Approaches. *Drug discovery today. Disease models 9*, e85–e90.

(*31*) Wang, D., Guo, Q., Wu, Z., Li, M., He, B., Du, Y., Zhang, K., and Tao, Y. (2024). Molecular mechanism of antihistamines recognition and regulation of the histamine H1 receptor. *Nature Communications 15*, 84.

(*32*) Komatsu, H. (2015). Novel Therapeutic GPCRs for Psychiatric Disorders. *International Journal of Molecular Sciences 16*, 14109–14121.

(*33*) Liu, Q., Garg, P., Hasdemir, B., Wang, L., Tuscano, E., Sever, E., Keane, E., Hernandez, A. G. L., Yuan, T. Z., Kwan, E., Lai, J., Szot, G., Paruthiyil, S., Axelrod, F., and Sato, A. K. (2021). Functional GLP-1R antibodies identified from a synthetic GPCR-focused library demonstrate potent blood glucose control. *mAbs 13*, Publisher: Taylor & Francis, DOI: 10.1080/19420862.2021.1893425.

(*34*) Milicevic, K., Rankovic, B., Andjus, P. R., Bataveljic, D., and Milovanovic, D. (2022). Emerging Roles for Phase Separation of RNA-Binding Proteins in Cellular Pathology of ALS. *Frontiers in Cell and Developmental Biology 10*, 840256.

(*35*) Dhanasekaran, R., Deutzmann, A., Mahauad-Fernandez, W. D., Hansen, A. S., Gouw, A. M., and Felsher, D. W. (2022). The MYC oncogene — the grand orchestrator of cancer growth and immune evasion. *Nature reviews. Clinical oncology 19*, 23–36.

(*36*) Hu, J., Pan, D., Li, G., Chen, K., and Hu, X. (2022). Regulation of programmed cell death by Brd4. *Cell Death & Disease 13*, Publisher: Nature Publishing Group, 1–15.

(*37*) Pei, X., Chen, Y., Liu, L., Meng, L., Zhang, J., Liu, Y., and Chen, L. (2024). E242-E261 region of MYC regulates liquid-liquid phase separation and tumor growth by providing negative charges. *The Journal of Biological Chemistry 300*, 107836.

(*38*) Yang, J., Chung, C.-I., Koach, J., Liu, H., Navalkar, A., He, H., Ma, Z., Zhao, Q., Yang, X., He, L., Mittag, T., Shen, Y., Weiss, W. A., and Shu, X. (2024). MYC phase separation selectively modulates the transcriptome. *Nature Structural & Molecular Biology*, Publisher: Nature Publishing Group, 1–13.

(*39*) Hill, T. L., *An Introduction to Statistical Thermodynamics*, New ed edition; Dover Publications Inc.: New York, 1987; 544 pp.

(*40*) Rubinstein, M., *Polymer Physics*, Google-Books-ID: EJrYoAEACAAJ; Oxford University Press: 2003; 440 pp.

(*41*) *Understanding Molecular Simulation*, 2023.

(*42*) Menczer, F., Fortunato, S., and Davis, C. A. A First Course in Network Science, Higher Education from Cambridge University Press ISBN: 9781108653947 Publisher: Cambridge University Press, `https://www.cambridge.org/highereducation/books/first-course-in-network-science/EE22722F27519D8BB1443C7225C57BAF` (accessed Oct. 15, 2024).

(*43*) Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1-2*, 19–25.

(*44*) Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845–854.

(*45*) Yoshizawa, T., Nozawa, R.-S., Jia, T. Z., Saio, T., and Mori, E. (2020). Biological phase separation: cell biology meets biophysics. *Biophysical Reviews 12*, 519–539.

(*46*) Flory, P. J. (1941). Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics 9*, 660.

(*47*) Huggins, M. L. (1941). Solutions of Long Chain Compounds. *The Journal of Chemical Physics 9*, 440.

(*48*) Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics 126*, 014101.

(*49*) Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., Schneider, R., and Bagos, P. G. (2011). Using graph theory to analyze biological networks. *BioData Mining 4*, 10.

(*50*) Kim, Y. S., and Hochstrasser, R. M. (2009). Applications of 2D IR Spectroscopy to Peptides, Proteins, and Hydrogen-Bond Dynamics. *The Journal of Physical Chemistry B 113*, Publisher: American Chemical Society, 8231–8251.

(*51*) Izadi, S., and Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *The Journal of Chemical Physics 145*, 074501.

(*52*) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics 81*, 3684–3690.

(*53*) Evans, D. J., and Holian, B. L. (1985). The Nose–Hoover thermostat. *The Journal of Chemical Physics 83*, 4069–4074.

(*54*) Parrinello, M., and Rahman, A. (1980). Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters 45*, Publisher: American Physical Society, 1196–1199.

(*55*) Maurya, M., Metya, A. K., Singh, J. K., and Saito, S. (2021). Effects of interfaces on structure and dynamics of water droplets on a graphene surface: A molecular dynamics study. *The Journal of Chemical Physics 154*, 164704.

(*56*) Mark, P., and Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A 105*, Publisher: American Chemical Society, 9954–9960.

(*57*) Krevert, C. S., Chavez, D., Chatterjee, S., Stelzl, L. S., Pütz, S., Roeters, S. J., Rudzinski, J. F., Fawzi, N. L., Girard, M., Parekh, S. H., and Hunger, J. (2023). Liquid–Liquid Phase Separation of the Intrinsically Disordered Domain of the Fused in Sarcoma Protein Results in Substantial Slowing of Hydration Dynamics. *The Journal of Physical Chemistry Letters 14*, 11224–11234.

(*58*) Roeters, S. J., Strunge, K., Pedersen, K. B., Golbek, T. W., Bregnhøj, M., Zhang, Y., Wang, Y., Dong, M., Nielsen, J., Otzen, D. E., Schiøtt, B., and Weidner, T. (2023). Elevated concentrations cause upright alpha-synuclein conformation at lipid interfaces. *Nature Communications 14*, Publisher: Nature Publishing Group, 5731.

(*59*) Murthy, A. C., Dignon, G. L., Kan, Y., Zerze, G. H., Parekh, S. H., Mittal, J., and Fawzi, N. L. (2019). Molecular interactions underlying liquidliquid phase separation of the FUS low-complexity domain. *Nature Structural & Molecular Biology 26*, Number: 7 Publisher: Nature Publishing Group, 637–648.

(*60*) Hamm, P., and Zanni, M., *Concepts and Methods of 2D Infrared Spectroscopy*, Google-Books-ID: hRbJ5n9bPGAC; Cambridge University Press: 2011; 297 pp.

(*61*) Ganim, Z., Chung, H. S., Smith, A. W., DeFlores, L. P., Jones, K. C., and Tokmakoff, A. (2008). Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Accounts of Chemical Research 41*, Publisher: American Chemical Society, 432–441.

(*62*) Woutersen, S., Pfister, R., Hamm, P., Mu, Y., Kosov, D. S., and Stock, G. (2002). Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. *The Journal of Chemical Physics 117*, 6833–6840.

(*63*) Torii, H. (2015). Amide I Vibrational Properties Affected by Hydrogen Bonding Out-of-Plane of the Peptide Group. *The Journal of Physical Chemistry Letters 6*, Publisher: American Chemical Society, 727–733.

(*64*) Buck, M., and Karplus, M. (2001). Hydrogen Bond Energetics: A Simulation and Statistical Analysis of N-Methyl Acetamide (NMA), Water, and Human Lysozyme. *The Journal of Physical Chemistry B 105*, Publisher: American Chemical Society, 11000–11015.

(*65*) Huerta-Viga, A., and Woutersen, S. (2013). Protein Denaturation with Guanidinium: A 2D-IR Study. *The Journal of Physical Chemistry Letters 4*, Publisher: American Chemical Society, 3397–3401.

(*66*) Krevert, C. S., Gunkel, L., Haese, C., and Hunger, J. (2022). Ion-specific binding of cations to the carboxylate and of anions to the amide of alanylalanine. *Communications Chemistry 5*, Number: 1 Publisher: Nature Publishing Group, 1–10.

(*67*) DeCamp, M. F., DeFlores, L., McCracken, J. M., Tokmakoff, A., Kwac, K., and Cho, M. (2005). Amide I Vibrational Dynamics of N-Methylacetamide in Polar Solvents: The Role of Electrostatic Interactions. *The Journal of Physical Chemistry B 109*, Publisher: American Chemical Society, 11016–11026.

(*68*) Middleton, C. T., Buchanan, L. E., Dunkelberger, E. B., and Zanni, M. T. (2011). Utilizing Lifetimes to Suppress Random Coil Features in 2D IR Spectra of Peptides. *The Journal of Physical Chemistry Letters 2*, Publisher: American Chemical Society, 2357–2361.

(*69*) Reid, K. M., Poudel, H., and Leitner, D. M. (2023). Dynamics of Hydrogen Bonds between Water and Intrinsically Disordered and Structured Regions of Proteins. *The Journal of Physical Chemistry B 127*, Publisher: American Chemical Society, 7839–7847.

(*70*) Schmidt-Engler, J. M., Zangl, R., Guldan, P., Morgner, N., and Bredenbeck, J. (2020). Exploring the 2D-IR repertoire of the –SCN label to study site-resolved dynamics and solvation in the calcium sensor protein calmodulin. *Physical Chemistry Chemical Physics 22*, Publisher: The Royal Society of Chemistry, 5463–5475.

(*71*) Ham, S., Kim, J.-H., Lee, H., and Cho, M. (2003). Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA–nD2O complexes. *The Journal of Chemical Physics 118*, 3491–3498.

(*72*) Kim, Y. S., and Hochstrasser, R. M. (2005). Chemical exchange 2D IR of hydrogen-bond making and breaking. *Proceedings of the National Academy of Sciences 102*, Publisher: Proceedings of the National Academy of Sciences, 11185–11190.

(*73*) Candelaresi, M., Ragnoni, E., Cappelli, C., Corozzi, A., Lima, M., Monti, S., Mennucci, B., Nuti, F., Papini, A. M., and Foggi, P. (2013). Conformational Analysis of Gly–Ala–NHMe in D2O and DMSO Solutions: A Two-Dimensional Infrared Spectroscopy Study. *The Journal of Physical Chemistry B 117*, Publisher: American Chemical Society, 14226–14237.

(*74*) Burke, K. A., Janke, A. M., Rhine, C. L., and Fawzi, N. L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular Cell 60*, Publisher: Elsevier, 231–241.

(*75*) Benayad, Z., von Bülow, S., Stelzl, L. S., and Hummer, G. (2021). Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. *Journal of Chemical Theory and Computation 17*, Publisher: American Chemical Society, 525–537.

(*76*) Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B 111*, Publisher: American Chemical Society, 7812–7824.

(*77*) Allouche, A.-R. (2011). Gabedit—A graphical user interface for computational chemistry softwares. *Journal of Computational Chemistry 32*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/ 174–182.

(*78*) Qi, Y., Ingólfsson, H. I., Cheng, X., Lee, J., Marrink, S. J., and Im, W. (2015). CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *Journal of Chemical Theory and Computation 11*, Publisher: American Chemical Society, 4486–4494.

(*79*) Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J., and Tieleman, D. P. (2014). Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation 10*, Publisher: American Chemical Society, 676–690.

(*80*) Robustelli, P., Piana, S., and Shaw, D. E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences 115*, Publisher: Proceedings of the National Academy of Sciences, E4758–E4766.

(*81*) Di Pierro, M., Elber, R., and Leimkuhler, B. (2015). A Stochastic Algorithm for the Isobaric–Isothermal Ensemble with Ewald Summations for All Long Range Forces. *Journal of Chemical Theory and Computation 11*, Publisher: American Chemical Society, 5624–5637.

(*82*) Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation 4*, Publisher: American Chemical Society, 116–122.

(*83*) Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., Domański, J., Dotson, D. L., Buchoux, S., Kenney, I. M., and Beckstein, O. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference*, Conference Name: Proceedings of the 15th Python in Science Conference, 98–105.

(*84*) Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry 32*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21787, 2319–2327.

(*85*) Gowers, R. J., and Carbone, P. (2015). A multiscale approach to model hydrogen bonding: The case of polyamide. *The Journal of Chemical Physics 142*, 224907.

Publication statement

The following publication resulted from this work:

Krevert, C.S., **Chavez, D.**, Chatterjee, S., Stelzl, L.S., Pütz, S., Roeters, S.J., Rudzinski, J.F., Fawzi, N.L., Girard, M., Parekh, S.H. and Hunger, J., 2023. Liquid–Liquid Phase Separation of the Intrinsically Disordered Domain of the Fused in Sarcoma Protein Results in Substantial Slowing of Hydration Dynamics. *The Journal of Physical Chemistry Letters*, 14(49), pp.11224-11234. DOI: `https://pubs.acs.org/doi/10.1021/acs.jpclett.3c02790`

Eidesstattliche Erklärung

I hereby declare that I have independently written this thesis and have only used the tools and resources explicitly stated. Specifically, I made use of the following tools to support my work:

- **ChatGPT:** For clarifying written ideas and refining formulations, without replacing my independent analysis and argumentation.
- **DeepL:** For translating abstract section to ensure clarity and precision in German language.
- **Grammarly:** For grammar and style corrections.

Daniel Chavez
Frankfurt am Main

Permission rights

Whenever a figure, table or text is identical to a previous publication, copyright permission and/or co-author agreement has been obtained as listed below.

The following parts of the thesis have been previously published by ACS Publications under Creative Commons Attribution License 4.0 (CC BY):

Krevert, C.S., **Chavez, D.**, Chatterjee, S., Stelzl, L.S., Pütz, S., Roeters, S.J., Rudzinski, J.F., Fawzi, N.L., Girard, M., Parekh, S.H. and Hunger, J., 2023. Liquid–Liquid Phase Separation of the Intrinsically Disordered Domain of the Fused in Sarcoma Protein Results in Substantial Slowing of Hydration Dynamics. *The Journal of Physical Chemistry Letters*, 14(49), pp.11224-11234.

- **Figure 3.10:** Experimental samples of FUS LC BC vs diluted phase.
- **Figure 3.11:** Linear expectra of FUS LC samples.
- **Figure 3.12:** Infrared spectra from MD simulations.
- **Figure 3.13:** 2D IR spectra of FUS LC.
- **Figure 3.14:** CLS lines for all samples.
- **Figure 3.16:** FUS clasification and time autocorrelation plot.

Image 1.1 **Protein Hierarchical organization** was created with BioRender.com and is used here in compliance with BioRender's academic licensing terms.
Created in BioRender. Chavez Rojas, D. (2025) https://BioRender.com/ l13k374
Agreement number HP27NVW9YD.

CV

Daniel Chavez Ph.D. Candidate

Contact Information:

- **Address:** Heinrich Seliger Str. 32, Frankfurt am Main
- **Email:** daniel109045@gmail.com
- **Phone:** +49 163 9633973

Education

- **Ph.D. in Sciences (Candidate)**: Max Planck Institute for Polymer Research, 2020–Ongoing
- **Master in Sciences, Physics**: University of Guadalajara, Mexico, 2016–2018
- **Bachelor in Physics**: University of Guadalajara, Mexico, 2010–2015

Research Experience

- **Ph.D. Project:** *The structural and dynamic characteristics of water in biological Systems*
  Max Planck Institute for Polymer Research (2020–Present)
- **Internship: ESCRT III Multiscale Simulations**
  Max Planck Institute for Biophysics (2019)
- **Master Thesis:** *Free energy perturbation analysis of Furin protein*
  University of Guadalajara and National Autonomous University of Mexico (2017–2018)
- **Bachelor Thesis:** *Biophysical study of opening and closing mechanisms in the ClC-2 Chloride Channel*
  University of Guadalajara and University of San Luis Potosi (2015–2016)

Publications and Conferences

- **Publication:** *J. Phys. Chem. Lett.*, 2023 Dec 14; 14(49): 11224–11234.

- Invited talk: *5th Material Sciences and Engineering Conference, University of Guadalajara*
- Presentations: DPG Spring Meeting (2022, 2023)

Workshops and Recognitions

- **Workshops:**
  - *51st Heidelberg Physics Graduate Days* – Generative learning for molecular systems (2023)
  - *CECAM Flagship Workshop: FAIR and TRUE Data Processing for Soft Matter Simulations* (2023)